U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

PB-252 641

TECHNICAL LITERATURE SEARCH SYSTEMS OF EPA

FINDINGS AND RECOMMENDATIONS

ARTHUR YOUNG AND COMPANY

PREPARED FOR
ENVIRONMENTAL PROTECTION AGENCY

JANUARY 1976

2

138186

PB 252 641

TECHNICAL LITERATURE SEARCH SYSTEMS OF EPA

FINDINGS AND RECOMMENDATIONS

TASK ORDER NO. 68-01-3077
CONTRACT NO. 68-01-1780

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

BIBLIOGRAPHIC DATA SHEET

1. Report No. EPA-220/1-76/001

2.

3. Recipient's Accession No.

4. Title and Subtitle
Technical Literature Search Systems of EPA
Findings and Recommendations

5. Report Date
January 1976

6.

7. Author(s)
Arthur Young and Company

8. Performing Organization Rept. No.

9. Performing Organization Name and Address

Arthur Young and Company
1025 Connecticut Avenue, N. W.
Washington, D. C. 20036

10. ~~Project~~/Task/~~Work Unit~~ No.
68-01-3077

11. Contract/~~Grant No.~~
68-01-1780

12. Sponsoring Organization Name and Address

Management Information & Data Systems Division
Environmental Protection Agency
401 M Street, S. W. Washington, D. C. 20460

13. Type of Report & Period Covered

14.

15. Supplementary Notes

16. Abstracts

The report contains a Management Summary and detailed findings and recommendations as to continued support of the technical information dissemination missions as it relates to the systems included in the study. The study evaluated the need for such systems and developed an action plan through which needed systems may be developed in an orderly fashion.

17. Key Words and Document Analysis. 17a. Descriptors

Bibliographic Systems	Abstract Systems
Technical Literature	Pollution Abstract Systems

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group

18. Availability Statement

Release Unlimited

19. Security Class (This Report)
UNCLASSIFIED

20. Security Class (This Page)
UNCLASSIFIED

21. No. of Pages 179

22. Price 7.50-2.25

FORM NTIS-35 (REV. 10-73) ENDORSED BY ANSI AND UNESCO. THIS FORM MAY BE REPRODUCED USCOMM-DC 8265-P74

ARTHUR YOUNG & COMPANY

1025 CONNECTICUT AVENUE N W
WASHINGTON D C 20036

February 2, 1976

Frank Bullock
Project Officer
Environmental Protection Agency
MIDSD (PM-218)
Washington, D.C. 20460

Reference: Contract No. 68-01-1780
Task Order No. 68-01-3077

Subject: Tasks 3 and 4 Deliverables

Dear Mr. Bullock:

Arthur Young & Company hereby submits the Tasks 3 and 4 deliverable products specified in the referenced Task Order. Provided is the FINDINGS AND RECOMMENDATIONS Report presenting the results of Tasks 3 and 4. The report contains a Management Summary, analyses of each of the EPA Technical Literature Search Systems, and the Recommendations and Action Plan for continued support of the technical information dissemination mission as it relates to the systems included in the study.

If you have any questions regarding this document please contact Mr. Robert F. Morse or myself at 202-785-4747

Very truly yours,

ARTHUR YOUNG & COMPANY

By: Hugo B Becker
Hugo B. Becker

TABLE OF CONTENTS

PAGE

INTRODUCTION 1

I. MANAGEMENT SUMMARY

1. OVERVIEW I-1
2. SUMMARY OF FINDINGS I-4
3. SUMMARY OF RECOMMENDATIONS I-7

II. SOLID WASTE INFORMATION RETRIEVAL SYSTEM II-1

III. NOISE INFORMATION SYSTEM III-1

IV. AIR POLLUTION TECHNICAL INFORMATION CENTER IV-1

V. PESTICIDES TECHNICAL/LIBRARY RESEARCH SYSTEM V-1

VI. EPA LIBRARY SYSTEMS VI-1

VII. RECOMMENDATIONS AND ACTION PLAN VII-1

1. PROBLEM SYMPTOMS AND RESULTS VII-2
2. MANAGEMENT RECOMMENDATIONS VII-8
3. OPERATIONAL RECOMMENDATIONS VII-10
4. ADDITIONAL CONSIDERATIONS VII-14

VIII. APPENDIX A - SAMPLE SEARCH ANALYSIS A-1

TABLE OF CONTENTS (Continued)

PAGE

IX. APPENDIX B - USER SURVEYS B-1

X. APPENDIX C - ANALYSIS OF REQUIRED JOURNALS C-1

XI. APPENDIX D - APTIC DATA PROCESSING PROBLEM LOG D-1

XII. APPENDIX E - COMPARISON SEARCHES E-1

XIII. APPENDIX F - ALTERNATIVES TO WYLBUR F-1

XIV. APPENDIX G - DATA BASE SYSTEMS AND MINICOMPUTERS G-1

LIST OF EXHIBITS

PAGE

I. LIBRARY SEARCH/SELECTIVE DISSEMINATION ACTIVITY VI-8

II. SWIRS-UTILIZATION/RESPONSE BY USER (EPA ONLY) A-3

III. SWIRS-UTILIZATION/RESPONSE BY USER (EXCLUDING EPA) A-4

IV. SWIRS-NUMBER OF SEARCHES BY GROUP A-5

V. SWIRS-AVERAGE TURNAROUND FOR SEARCHES A-6

VI. SWIRS-SOURCE OF KNOWLEDGE OF SWIRS AS INDICATED BY REQUESTS A-6

VII. NOISE-USER UTILIZATION (EPA ONLY) A-7

VIII. NOISE-USER UTILIZATION (EXCLUDING EPA) A-8

IX. APTIC-USER UTILIZATION (EPA ONLY) A-9

X. APTIC-USER UTILIZATION (EXCLUDING EPA) A-10

XI. APTIC-UTILIZATION BY SUBJECT (EPA ONLY) A-12

XII. APTIC-UTILIZATION BY SUBJECT (EXCLUDING EPA) A-13

XIII. PESTICIDES-AVERAGE HITS BY DATA BASE SEARCHED A-15

XIV. INTERVIEW FORMS-TECHNICAL LITERATURE SEARCH SYSTEMS B-10

XV. NUMBER OF HITS BY DATA BASE E-3

XVI. DATA BASE MANAGEMENT SYSTEMS COMPARISON G-9

INTRODUCTION

This FINDINGS AND RECOMMENDATIONS REPORT is the culmination of the Study to Analyze the Technical Literature Search Systems of EPA conducted by Arthur Young & Company under the Task Order 68-01-3077.

The stated purpose of the study was to

> "Perform an analysis of the Agency's Technical Literature Systems; to determine the need for such systems and to develop an action plan through which needed systems may be developed in an orderly fashion."

As a part of the study Arthur Young and Company was directed to

- Review the physical characteristics of the four program related Bibliographic Systems and the Library Systems in order to document the characteristics in the form of of SYSTEMS PROFILES delivered as a product of Task 1 of the study
- Determine management requirements, plans and objectives for the future of the Literature Search Systems and present them as MANAGEMENT OBJECTIVES delivered as a product of Task 2 of the study
- Develop feasible approaches to the support of the EPA Literature Search Systems as incorporated in this report
- Prepare recommendations and an action plan for continued support of the technical literature dissemination mission of EPA as it relates to the systems included in the study

In meeting the objectives of the study the Arthur Young and Company Project Team examined the technical literature usage and dissemination policy of all systems as a whole and the characteristics, costs and mission of each object system individually.

The systems included in the study were

- Solid Waste Information Retrieval System (SWIRS)
- Noise Information Service (NOISE)
- Air Pollution Technical Information Center (APTIC)

I. MANAGEMENT SUMMARY

1. OVERVIEW

The most significant observation resulting from this study is that the Technical Literature Search Systems of EPA generally lack effective and continuous management direction and control by which the EPA might assure achievement of the mission and objectives of those systems.

The Arthur Young and Company Project Team has, almost uniformly, found:

- No system justification statements or performance requirements
- No EPA or program policy statements regarding information dissemination
- No program management objectives for system operations or missions.

With the exception of EPA Orders to support the Library operations, only vague references to general Federal Acts can justify systems operations. The symptoms and effects of this lack of management control have been apparent throughout this study project. Among the most serious are:

- High cost in comparison to the level of services rendered
- Very limited communication and coordination between individual systems, with other EPA program areas or with other Federal Agencies
- Lack of effective communication and assistance from EPA service offices, such as the Office of Public Affairs and the Management Information and Data Systems Division
- Continued development of myriad systems performing the same general activity.

It would be unjust to place the primary responsibility for system inadequacies upon the System Managers or even middle management. The System Managers have performed to the highest level which can be reasonably expected. They have been required by the EPA to be experts as librarians, public relations personnel, data systems analysts and financial managers. The lack of a definition of scope of these activities has caused them to attempt to provide an optimum range of benefits utilizing a limited budget: data bases and libraries containing all pertinent, non-duplicative literature; complete query and searching support without charge; and, frequently, advertisement of the systems to all potential users. In brief, they are attempting more than their resources permit.

The Systems Managers have had little management direction and few expectations and objectives to be met for their management. Directives have been primarily negative in the form of budgetary and personnel reductions without corresponding reductions or changes in defined scope of system coverage.

Without management direction and planning, the Systems Managers have tended to isolate themselves and their function and have not been forthcoming with suggestions for changes in functional direction. As a result, even if budgets were expanded, without corresponding management direction, system expansion may not be toward improved services but would likely be toward expanding existing misdirected efforts.

Without the definition and enforcement of policies or guidelines requiring in-depth research, planning and reporting, the systems have come to exhibit the following characteristics:

- Limited System Usage - None of the four primary Technical Literature Systems (SWIRS, APTIC, NOISE, PESTICIDES) enjoy any reasonable level of usage, with the result that cost/usage figures are extremely high. Usage data is not available for the EPA Library Systems

- Limited Cost Data - Few of the services were able to provide any reasonable amount of cost data, and none apparently performed any cost-effective analysis on the systems

- Insufficient Knowledge of Usage and Holdings - Only the APTIC System can provide a breakdown of holdings in terms of subject, level of specificity and source. None of the systems retains information on items retrieved, so that none of the services can analyze the utilization of its data base.

- Lack of Data Processing Coordination - Although only the new NOISE System Manager possesses appreciable experience and expertise in the field of data processing, each of the Systems Managers has independly developed and maintained the computer system supporting each of the Technical Information Systems. There is little or no communication with the Management Information and Data Systems Division, and data processing feasibility studies were evidently not performed prior to system development.

- Lack of Coordination Among System Managers - While the individual systems each contain information which would be valuable to other systems, such as the EPA Library holdings of legislative and regulatory data, the Technical Literature Services do not utilize each other's holdings, either in collecting information or in satisfying search requests.

- Lack of Cooperation with other EPA Offices or other Federal Agencies - The potential for cooperative effort with other information services, such as the office of Research and Development's Technology Transfer Program, the Office of Public Affairs, the National Library of Medicine or the Energy Research and Development Agency, is not being realized.

- Inappropriate Allocation of Effort - The EPA Technical Information Services are suffering from an inappropriate allocation of funds and attention. The majority of effort expended by these services concentrates on the development of comprehensive data bases and the response to requests with the most inclusive product possible. Only minimal effort is expended in publicizing the system and educating potential users. While this approach is appropriate for a well-known, heavily-utilized system, it is the opposite of the requirements of these systems. The excellence of a library or the completeness of a data base is of little importance if a system is not used, and the use of the EPA Technical Literature Systems is minimal.

2. SUMMARY OF FINDINGS

The following discussion presents brief summaries of the most important characteristics and critical problems noted during the analysis of each of the EPA Technical Information Systems. These characteristics, and other findings and recommendations of the Project Team, are discussed in detail in Chapters II-VI.

(1) Solid Waste Information Retrieval System

A great deal of effort has gone into the creation of this information source which apparently can effectively respond to the vast majority of queries on the subject of Solid Waste Management. However, this source is not only very expensive to maintain, but its utilization is very low. The Project Team is not certain that the data processing system could economically support a sizeable increase in number of requests. If the Office of Solid Waste Management Programs is to fulfill the requirements of the Solid Waste and Disposal Act as it has interpreted them, it must take steps and incur expense, which will increase the level of recognition and usage of SWIRS services. Since OSWMP cannot increase its expenditures for SWIRS (and in fact desires to reduce them), the funds for any increase in publicity and training must come from a reduction in other SWIRS activities. While the development of a more efficient data processing system and the elimination of WYLBUR would reduce total costs, it would be unrealistic to anticipate a savings greater than $50,000 from these actions. Additional reduction in expenditure must, therefore, come from the selection, acquisitioning and abstracting effort to support costs expended in increasing system usage.

(2) NOISE Information Service

The NOISE Information Service appears to be consistently underutilized. While users express satisfcation with service provided on specific limited subjects, they consider the overall scope of subject coverage to be inadequate. Changes in the direction of this program, due to a transfer of funding operational responsibility, are being planned, with selection activity to be performed by the Office of Noise Abatement and Control (ONAC) as a part of its normal operations. This should substantially reduce current expenditures for contractor support.

ONAC must determine the extent of external user need, and the extent to which it intends to fill this need. ONAC Program Management intends to perform an analysis of ONAC's own information requirements and to modify the functions of NOISE to fill these requirements. The system can be expected to become a combination bibliographic and automated filing system. Any system development activity should be performed in cooperation with the Management Information and Data Systems Division, in order to create a system which will not only fill the needs of ONAC, but which could also function as a generalized bibliographic system capable of supporting any of the EPA Technical Information Systems.

(3) Air Pollution Technical Information Center

It would seem that there would be a continuing National interest in an air pollution information retrieval system. However, actual usage of APTIC peaked in 1972 and has decreased by 50% since that time. With a current annual cost of $723,108 and annual usage of 876 requests, the justification for maintenance of APTIC at current (or even 1972) usage levels is questionable. While part of this decline may be due to the serious data processing problems APTIC is experiencing, it is difficult to ascribe all of the decline to slow turnaround, particularly since the decline began before the implementation of the new system. Finally, while APTIC strives to serve those who cannot acquire such information elsewhere, only a small minority of current users fall into this category.

If the Air Pollution Technical Information Center is to fulfill the mission of providing information to all who need it, it must vastly increase the number of users benefiting from the system, particularly those without an ability to acquire this data elsewhere. If the amount of effective usage of the system cannot be increased there is little justification for its continued existence. Since the funds and effort needed to increase usage of the system must come from elsewhere in the APTIC service, it seems unavoidable that acquisition and/or data processing expenditures must be reduced.

(4) Pesticides Technical/Library Research System

The Pesticides System is, in terms of cost per request serviced, the most expensive of all the EPA Technical Literature Systems, thought it does not expend the greatest total amount. It is impossible to estimate the impact of the upcoming registration effort to be performed by the Office of Pesticide Programs (OPP), particularly since no detailed determination of this activity's information needs has been made. It is assumed by OPP that a great deal of detailed information will be required. Current search activity is in response to extremely general queries frequently resulting in the retrieval of hundreds of abstracts. This is in part due to the practice by the OPP Criteria and Evaluation (C & E) Division of defining all of its own search requirements through its Information Office. C & E curren[illegible] is also performing and funding half of its searches utilizing commercial data bases.

(5) EPA Library Systems

The three information systems directed by the EPA Library itself (EPA Reports, International Reports and Document Control Systems) are actually not literature search systems but integral parts of library operations, acting as automated catalogs and supporting inventory activities. As such they are not candidates for elimination. While the design employed for these systems is simple and effective, they should not have been developed as separate systems utilizing a single concept, but as a single integrated system. This concept of a single, integrated technical literature system should be considered by the EPA Library in any future system developments. The library is designing its own systems, through its own contractor, and is not receiving the benefits of expertise within MIDSD.

The Air Ecology System appears to operate in isolation from the rest of the EPA, even though it provides microfiche copies of citations to the various EPA activities. While user statistics are not available, the total annual cost of the system is only approximately $45,000. Most of the literature entering the system is selected from secondary sources rather than the result of an independent review of source journals.

3. SUMMARY OF RECOMMENDATIONS

The following recommendations are presented in order that the Environmental Protection Agency may effectively develop and coordinate the activities of the EPA Technical Literature Search Systems in providing needed literature to its chosen user population. These recommendations have been developed to enable each of the EPA Technical Literature Search Operations to more effectively and economically serve their chosen user population, and to assist the EPA in ensuring overall efficiency in the dissemination of subject literature.

(1) EPA Must Establish A Comprehensive And Uniform Information Dissemination Policy

There is a critical need at the Environmental Protection Agency for an inclusive, Agency-wide policy concerning the dissemination of information regarding the causes and control of pollution. The individual Acts and

Regulations referenced by the Technical Literature Services do not appear to either mandate or preclude the type of information systems in operation at EPA. Rather, these systems have come into being either on the basis of decisions made at the middle management level or because of the assignment to EPA of other governmental units already supporting such a system. Consequently, before particular decisions can be made regarding the continued existence or characteristics of each of the systems, the mission and requirements of the service must be verified or redeveloped in concert with an overall policy on information dissemination. Since such a policy does not currently exist, it is incumbent upon the EPA to examine its own mission in this light, develop this policy and establish the proper working relationships between the EPA Technical Information Services, the EPA Library, the Office of Public Affairs, the Office of Research and Development and other EPA operating units.

While the development of an EPA Information Dissemination Policy is not within the scope of this project, the primary questions to be addressed in this effort are depicted in the RECOMMENDATIONS AND ACTION PLAN.

(2) A Single Central Technical Literature Coordinator Is Needed To Plan And Monitor All Systems Activities

It is recommended that an individual be assigned the position of Technical Literature Coordinator, to advise, coordinate and review the activities of each of the System Managers. The coordinator would provide them with the expertise in financial management, program planning, user training and public relations activities which the systems currently are lacking. It is not recommended that the Technical Literature Services be transferred from their current organizational position since they either support program needs or operate in tandem with other program activities. The Technical Literature Coordinator might occupy a staff position within the Office of Planning and Management, and would devote his full time to this position. His qualifications would include substantial ability and experience in financial management and control, statistical analysis and program planning and development. In this position he would direct the individual system managers in implementing the recommendations contained in this report and in developing improved methods of providing service to system users.

A Data Processing Specialist should be assigned to assist the Technical Literature Coordinator in any activities requiring computer expertise. This specialist should have experience in data base management systems and in input equipment selection. While the extent of his participation will depend upon the decisions made regarding the disposition of the systems, the Technical Literature Systems must be his primary responsibility, so that he can devote whatever time is necessary without significant interruption from other duties.

(3) EPA Should Consider Commercial Search Services As A Viable Alternative To Current Systems Operations

While it may be impossible to place a value upon timely, pertinent information, Arthur Young & Company considers the search costs of the systems to be excessive, particularly since commercial services are available which may provide comparable information at substantially lower costs. As demonstrated in the COMMERCIAL ABSTRACTING SERVICES analyses for each of the systems, savings from $183,000 (Pesticides) to $516,700(APTIC) would result from a discontinuation of selection and acquisition activities and a reliance upon commercial services in fulfilling requests for information. These analyses also demonstrate that, even if it is necessary for the Technical Literature Services to continue selecting and abstracting at half their current level, donating the product to commercial services to ensure their inclusion, considerable savings still result. The recommendation for the use of commercial services does not include the three library systems, but does include the Air Ecology System. NOISE would not be included under its planned concept of retaining primarily internally generated documents, but should consider the use of commercial sources in any searches of externally generated literature.

The primary advantages offered by commercial services include:

- Substantial decrease in operating cost
- Immediate improvement in SWIRS and APTIC service, since turnaround problems due to obsolete computer system design (SWIRS) and system problems and down time (APTIC) would be bypassed
- Release of personnel time and funds to increase the recognition and sophistication of potential users, improve search quality, and perform the currently neg-

lected management duties discussed above.

Each Technical Literature Service must verify the ability of commercial data bases to properly respond to its own requirements. The following recommendations list in chronological sequence, the steps which should be taken in evaluating the completeness and usefulness of the commercial data bases and, if necessary, the actions which might be taken to improve the expected performance from the commercial services.

(4) EPA Must Review Commercial Data Base Coverage To Assure Search Adequacy

While a sizeable overlap exists between the EPA Technical Literature Systems and the major commercial systems in terms of journals reviewed, it cannot be guaranteed that this same overlap exists in terms of articles selected. Therefore, for each of the systems being considered for replacement, a computer listing of the last 500 selections sorted by author, should be generated. This listing could then be compared against a similar listing of the major commercial sources. If the vast majority of articles present on the EPA System are present on the commercial system(s), EPA should begin immediately to discontinue its own system and utilize the commercial systems.

(5) Donation or Sale of EPA Data Bases to Commercial Services Would Insure Complete Coverage

If no commercial data base(s) is discovered with a sufficient amount of duplicated selections with the corresponding EPA Data Base, EPA should attempt to sell or even donate its current data base to one or more of the commercial abstracting services. It is quite possible that one or more of the commercial abstracting services would agree to expand their coverage to include most EPA-related items, in return for the usage of the associated EPA data base. Even if such an agreement cannot be reached, it would be substantially cheaper for EPA to continue its selection and acquisition activity at ½ the current level, donate the results to a commercial service and use that service to search the overall data base.

(6) If Commercial Service Data Bases Are Unacceptable EPA Should Maintain a Single Technical Information Data Base

Only if each of the above approaches fail would

Arthur Young & Company recommend the continuation of the Pesticides, SWIRS or APTIC data bases, and then under these conditions:

- EPA should purchase abstract tapes from commercial services rather than contracting the abstracting activity, for those journals and subject areas effectively covered by the commercial services.

- The resultant EPA data base should be run not on an EPA system but on one of the commercial services, such as SDC or Lockheed. These services are interested in purchasing the EPA data bases and servicing them just as a commercial data base. They would pay EPA a fixed fee and royalty, while EPA would pay a reduced rate to access its own data base.

(7) If EPA Operations of the Literature Search Systems is Essential a Redesign of Current Systems is Necessary

Should the EPA decide to retain its own data base for one or more of the Technical Literature Search Systems, it is not recommended that it utilize its own data processing system for any data base which it is able to put up on a commercial data processing system. However, if either SWIRS or APTIC is retained as a data processing system, then serious consideration should be given to designing a new data processing system providing on-line and batch searching using inverted files. This is recommended since both SWIRS and APTIC currently possess serious problems in terms of data processing performance and would cost more to upgrade than to redesign, particularly if an efficient data base management system is utilized. If the Pesticides System is retained, it could be left just as it is, as long as almost all searching is performed on microfiche. If frequent computerized searching should be begun for Pesticides, it should be redesigned along with SWIRS or APTIC.

II. SOLID WASTE INFORMATION RETRIEVAL SYSTEM

The Solid Waste Information Retrieval System (SWIRS) exists to provide for dissemination of literature regarding solid waste management to as wide an audience as possible, in order to fulfill the provisions of the Solid Waste and Disposal Act of 1965. This act requires the EPA to encourage, cooperate and assist public and private institutions and organizations, and private individuals in the conduct and coordination of research, investigations, experiments, training, demonstrations, surveys and studies in the field of solid waste management. The bill authorizes the EPA to collect and make available, through publications and other appropriate means, information regarding this subject area. Through SWIRS, and through other publication and promotional activities, the Technical Information Staff, Office of Solid Waste Management Programs (OSWMP) attempts to educate all segments of society in this area and to provide a useful and timely response to queries for assistance in locating and acquiring literature regarding Solid Waste Management.

1. SYSTEM DESCRIPTION

This System Description presents a brief view of the manner in which SWIRS strives to fulfill its responsibilities under the Solid Waste and Disposal Act as interpreted by OSWMP. A more detailed description may be found in the "EPA Technical Literature Systems Profiles," prepared separately under this Task Order.

Since OSWMP currently performs no regulatory functions, it must rely upon educational activities to increase awareness of the problems associated with solid waste management and their potential solution. Since most solid waste activity occurs in the local sector, OSWMP concentrates on state and local government, private enterprise, citizen action groups and the public in disseminating information. SWIRS Management is particularly concerned that information be provided to those persons not able to acquire this information elsewhere, and is therefore reluctant to demand any reimbursement for their services.

SWIRS is currently operated under contract by the Franklin Institute in Rockville, Maryland, although this contractor activity will shortly be re-bid. The contractor screens the world's literature for articles relating to solid waste management, selects copies and abstracts those which should be included in the SWIRS library and computer data base, updates the data base with the citation and abstract, and maintains the library documents selected for inclusion in the system.

SWIRS designates a set of 700 "core journals" which the contractor must screen; the contractor is free to screen any other journal which he believes will be productive. The contractor is expected to select, acquire and abstract approximately 6,000 documents annually.

The contractor is also responsible for responding to all requests for information and assistance. EPA requests for services are collected at the EPA and delivered to the contractor. Outside requests are mailed to a Post Office box number for pickup by the contractor. Each request is assigned a search number, logged in, and reviewed to determine the proper response materials, using a preprinted checklist. The search analyst must determine whether a "canned" search (a previously designed and executed search) or a new computer search is needed. If a similar query has been previously processed, the analyst shall use the results of that search, and only search the data base itself for citations entered after the date of the previous search. If possible, the results of the computer search are screened to ensure relevancy; this is currently only being performed for EPA requests. The requestor is mailed abstracts, on either computer printout or microfilm reader/printer copy.

Several other sources may be examined in fulfilling these requests. The Solid Waste Management Available Information Materials Manual contains listings of both OSWMP publications and NTIS publications. SWIRS orders relevant OSWMP publications, in the name of the requestor, from Cincinnati and provides the requestor with copies of the NTIS list, annotating relevant citations. Xerox copies are also provided of relevant pages from:

- Summaries of Solid Waste Intramural Research and Development Grant Projects (1971)
- Solid Waste Demonstration Grant Projects (1971)
- Summaries of Solid Waste Research and Training Grants (1970)
- Abstracts, Excerpts, and Reviews of the Solid Waste Literature (1972)
- Summary of Solid Waste Management Contracts (1975)
- Bureau of Mines Information Circular #895, "Research Programs on Recycling and Disposal of Mineral -, Metal -, and Energy - Based Wastes" (1973)

The SWIRS library provides copies of requested articles to EPA users, and will provide some inter-library loan services. However, SWIRS prefers that non-EPA users acquire copies from public libraries and other facilities whenever possible.

The SWIRS data processing system is a sequential batch system operated at Optimum Systems, Incorporated (OSI) in Rockville, Maryland through typewriter - type terminals located at the contractor facility. New citations and abstracts are entered to temporary files via WYLBUR, a text-editor, and later input to an edit program. After correction of any errors, the work files are merged in Accession Number sequence and then merged with the SWIRS Master File. All searches are performed by loading boolean search strategies onto temporary files (one search strategy per file) and inputting those files into the search program, which sequentially searches one or more of the master file tape volumes, depending on the publication dates over which the data base is to be researched. Program output can be either Accession Numbers only or entire abstracts of selected citations. Accession Number output is used when the search results are to be screened. This involves locating and reviewing the abstract on microfilm and, if the abstract is selected, generating a copy of the abstract on the microfilm reader/printer. Because of personnel costs, screening has been eliminated for all but EPA-generated searches; abstracts are mailed directly as printed by the computer.

2. MANAGEMENT PLANS AND OBJECTIVES

The Management Plans, Needs and Objectives of the Office of Solid Waste Management Programs supported by SWIRS have been documented in the "PROGRAM MANAGEMENT OBJECTIVES REPORT" produced on December 15, 1975 under this Task Order. The primary factors noted by management are discussed below:

(1) Program Management

While pleased with the overall performance of SWIRS, management hopes that expenses can be reduced without serious degradation of performance. However, without detailed management objectives, performance requirements and cost/benefit analyses, management is not really able to make judgments regarding performance. SWIRS does provide considerably more cost, usage and growth documentation than the other technical literature systems, including user surveys, weekly activity reports and internally generated description of costs, services and growth. Program Management should review this information against previously determined goals and assist SWIRS in improving their data gathering and analysis.

(2) Contractor Monitoring

Management believes that the use of contract services for all SWIRS functions, made necessary by program funding limitations, does not allow sufficient operating flexibility and control. In addition, the many responsibilities of the System Manager, including contract operations review, selection of nonperiodicals and procurement of OSWMP subscriptions and documents, do not permit him to publicize the system or educate users to the extent necessary. Arthur Young & Company concurs with this opinion and believes that this limitation not only prevents proper publicity of the system, but also reduces the opportunity for proper management of the system. The System Manager cannot properly monitor the contractor's performance, and is highly dependent upon the contractor for both performance and research. This is exemplified by the following:

- Both the SWIRS User Surveys and a review of commercial abstracting services were performed not by the EPA but by Franklin Institute, the contractor whose performance was in part being evaluated by the studies, and who certainly had a monetary interest in their results. While the contractor cannot be expected to refuse to perform these services, such activities should be performed by OSWMP or a third party.

- Although the "Summary of the Joint Meeting of APTIC, HAPAB and SWIRS Project Officers with common Contractor," December 10, 1973, states that there is less than 6 percent overlap among required journal titles, the analysis performed by the Project Team indicates a much higher percentage of overlap, such as the 35.3% of APTIC also reviewed by SWIRS and the 25.0% of Pesticides reviewed by SWIRS. These findings are presented in Appendix C, the ANALYSIS OF REQUIRED JOURNALS.

- Although the contractor is required to respond to every single search request within 7 days, the average turnaround time, as indicated by the Project Team SAMPLE SEARCH ANALYSIS (Appendix A), is 10.7 days. This was discovered by the System Manager some time ago, who demanded and was assured that the 7 day turnaround requirement would be met. However, of the fifty-nine requests processed during October, 1975, only six were finished in seven days or less. Three requests were from EPA users and none of these were processed within seven days.

- While contract search personnel are required to use discretion in responding to queries with an amount and type of literature responsive to the apparent need, the project team was often able to discover little difference in responses to very general or even superficial requests; e.g., send me literature about recycling, and to quite specific queries. This problem is addressed in the discussion on User Service in the Analysis Observed characteristics and in Appendix A, the SAMPLE SEARCH ANALYSIS.

(3) Information Management

Program Management recognizes that the field of Solid Waste Management is growing at a tremendous rate, and that new subjects should be included in the scope of covered literature. In addition, management anticipates the assignment of regulatory responsibility, with a corresponding increase in information needs. This presents a problem of finances since, without increased funds, SWIRS cannot in its present configuration increase its subject coverage.

3. ANALYSIS OF OBSERVED CHARACTERISTICS

This discussion delineates those characteristics of SWIRS operations which the project team considered to be most germane to the purposes of this study or in which particular improvements can be made. Where appropriate, comments are made concerning the impact of particular factors upon overall system performance. The characteristics reviewed during this section are also used in supporting discussions and recommendations written in the PRIMARY CONSIDERATIONS Section and the RECOMMENDATIONS AND ACTION PLAN.

(1) System Usage

SWIRS currently supports approximately 1,200 searches annually, which are developed from approximately 624 requests for information (a single request may cause several searches to be run; one for each distinct question or subject contained in the request). These figures were developed from an examination of documentation of actual searches made during five particular months (November, 1974; May 1975; August through October, 1975). The five month figures were then extrapolated to reflect annual values. The Project Team SAMPLE SEARCH ANALYSIS (described in Appendix A) and the USER SURVEYS (Appendix B) are the primary sources of usage information in this study. All figures quoted in this section have been extrapolated to reflect annual values. Of the 1,200 searches, 116 were 'canned', and used the results of previous searches in their response, rather than requiring a computer search. System Management's own figures, developed by SWIRS Operations by counting and extrapolating the number of searches for the first nine months of 1975, reflect an annual number of searches of 919 (SWIRS Operations did not tabulate the number of requests).

The estimated level of System usage, by type of user, is as follows:

USER GROUP		NUMBER OF SEARCHES		
		PROJECT TEAM ANALYSIS		SWIRS OPERATIONS ANALYSIS
EPA*		110 (9.2%)		128 (12.9%)
OSWMP			69 (7.5%)	
OTHER EPA			59 (6.4%)	
INDUSTRY/COMMERCIAL		379 (31.6%)		273 (29.7%)
FEDERAL GOVERNMENT		58 (4.8%)		48 (5.2%)
STATE/LOCAL GOVERNMENT		149 (12.4%)		132 (14.7%)
ACADEMIC		170 (14.2%)		138 (15.0%)
OTHER**		334 (27.8%)		200 (21.8%)
PRIVATE CITIZENS	209 (17.4%)			
STUDENTS	60 (5.0%)			
CITIZEN ACTION GROUPS	24 (2.0%)			
FOREIGN	41 (3.4%)			
TOTAL		1,200 (100%)		919 (100%)

* The Project Team analysis did not separate EPA users by OSWMP and other

** The SWIRS operations analysis did not separate 'other' users into its components.

These figures indicate that private industry represents the single largest group of users, with 379 annual searches, or 31.8% of the total. While EPA contractors are included in this group, the number of commercial requests identifying themselves as contractors represents approximately 7% of commercial requests. As expected, EPA usage of SWIRS is limited, although the System Manager himself does respond to a number of additional EPA requests by providing copies of books and reports, from his own collection of reports, publications and manuscripts at EPA. The usage of the system by Private Citizens and Citizen Action Groups is extremely limited, while usage by students (5.0%), academia (14.2%) and other Government (17.2%) represents a substantial proportion of all users. A portion of the requests submitted by academia were for projects being externally funded under the auspices of the academic institution. The Project Team did note that, over the past year, there has been an increase in the number of requests from industry and a decrease in those from Private Citizens. System Management believes that this may be a result of changes in the type of publicity which SWIRS has received. There were apparently more articles in general purpose and ecology publications last year concerning SWIRS than this year. The reason for the increase in commercial users is not clear, but may reflect a growing interest by these groups or a growing pressure upon them for positive

In performing the analysis of searches performed, described in Appendices A and B, the Project Team categorized information requests into the following three levels of detail:

- General, including requests for information covering a broad subject without defining particular methodologies such as "Solid Waste Management" or 'Municipal Waste Disposal'

- Semi-Detailed, which identifies an area within a larger category such as Municipal Waste Disposal-Incineration

- Detailed, which specifically identifies a particular aspect of a subject area. For example; "Municipal Waste Disposal-Incineration; Von Roll Process."

The following table exhibits the level of detail which was requested by the various user groups, extrapolated to represent the number of annual queries:

	Number of Searches by User Group		
User Group	General	Semidetailed	Detailed
EPA	5	34	72
Industry/Commercial	10	204	166
Federal Government	2	29	26
State/Local Government	12	91	46
Academic	17	89	65
Students	5	34	22
Private Citizens	70	96	43
Foreign	2	12	26
Civic Groups	2	17	5
Total	125	606	471

The Project Team was able to make the following general observations, based on the user interviews and search analyses:

- Federal Government and Commercial queries were normally quite specific, including the purpose for the information desired. Other users were somewhat less specific and did not normally provide the use to which the information would be put

- A relatively large number of Private Citizens wanted information regarding such subject as the disposal of human waste in rural areas and the use of compost and fertilizer. Few Private Citizens submitted queries regarding community oriented issues.

- Academia, composed of college teachers and researchers, were normally engaged in some project or study, often funded externally rather than by the college.

- Students requested information only to support a class project, term paper, or thesis. In several instances the search material was needed by a certain date. Frequently the date was not met by the contractor as most requests wanted the information within 2-3 weeks. The student group comprised only 5% of the total requests.

(2) Operations Costs

The cost and budgetary information presented in this section refers to Fiscal Year 1975, since this is the period for which the most complete and reliable figures could be obtained. While these figures will be slightly different from those costs to be experienced over the next year, this difference should not be substantial, with one probable exception. Computer operations costs may be expected to be moderately to substantially higher at the current vendor, Optimium Systems Incorporated, than they were at the National Institutes of Health. The costs incurred by SWIRS during Fiscal Year 1975 are as follows:

Service		Expenditure
Information Retrieval Services Branch-EPA		$ 33,327
NTIS and GPO Interagency Agreements		800
Franklin Institute, Total		362,414
Screening, Acquisitioning, etc. (7080 documents)	$ 219,312	
Searches (1440 searches)	81,988	
Library Cataloging (1800 documents)	16,480	
SWIRS Library Maintenance	16,333	
Programming, merge	4,090	
Translations	24,211	
NIH Computer		116,769
OSI Computer		15,208
TOTAL		$528,518

The figures for Franklin Institute were developed by prorating the total Franklin invoices for 1975 among the costs and services described in the April/June cost Proposal. (That cost proposal reflected an annual charge of $ 356,260 rather than the actual billing of $ 362,414.)

These figures were used to develop the overall cost of the system per search, and the costs which could be anticipated if the number of searches performed were to be tripled, due to growth in system usage. Several estimates of request and search activity were used in computing cost/search and cost/request:

- 1200 annual searches, estimated by Arthur Young Sample Search Analysis (A)
- 1440 annual searches, estimated by Franklin Institute April/June Cost Proposal (B)
- 624 annual requests, estimated by Arthur Young Sample Search Analysis (C)

	COST/SEARCH		COST/REQUEST
Cost Component	A(1200)	B(1440)	C(624)
VARIABLE COSTS			
Contractor Search Charge (81,988)	$ 68.32	$ 56.93	$ 131.39
Data Processing; 1/3 NIH Charge (38,923)	32.44	27.03	62.38
FIXED COSTS			
Contractor Fixed Charge (239,882)	199.90	166.58	384.43
Data Processing Fixed Charge (77,846)	64.87	54.06	124.75
TOTAL - All Costs ($438,639)	$365.53	$304.60	$702.95

The calculations do not include expenses allocated to the Information Retrieval Services Branch, Interagency Agreements, maintenance of the SWIRS Library or contracted translation services (a total of $ 74,671) since these activities are not critical to the satisfaction of search requests. In addition, OSI computer charges of $ 15,208 are omitted, since this may be primarily attributable to system conversion. Because of the high cost of WYLBUR entry, the Project Team has allocated only 1/3 of data processing costs to search activity, with other 2/3 being allocated to data base creation and maintenance.

In order to estimate the costs which would be experienced if system usage were to triple (thus lowering per/search cost), the following figures have been developed. They represent a tripling of number of searches and corresponding variable cost, with constant fixed cost.

		COST/SEARCH		COST/REQUEST
Cost Component		A(1200x3= 3600	B(1440x3= 4320)	C(624x3= 1872
Variable Cost (120,911x3) =	362,733	100.76	83.97	193.76
Fixed Cost =	317,728	88.26	73.55	169.73
Total All Cost	680,461	189.02	157.52	363.49

Even the most optimistic of these cost figures reflect a current cost of $304 search and future cost (with number of searches tripled) of $ 157/search. The Project team has noted that the cost proposal for the latest 3 month contract extension (October - December) reflects a contractor cost/search of $ 35, rather than the Fiscal Year '75 figure of $ 55. Presumably this decrease reflects an increase in contractor competence and efficiency in responding to queries. However, the number of documents to be abstracted has increased for this period, with the result that the total cost for the quarter is $ 3,865 higher than an average quarter of Fiscal Year 1975. If this pattern continues, it represents a decrease in variable costs/search and an increase in fixed costs, with no change in current cost/search. With increased growth, cost/search would be somewhat less. However, even if operations cost were to be reduced by $20/search, with no increase in fixed cost, cost per search would still be extremely expensive ($284/search currently and $137/search with a tripling of usage). Finally, it must be noted these figures represent searches, not individual requests. As indicated above, current cost per request is $703.

(3) System Service

The overall service provided by SWIRS appears to be responsive to the needs of its users. According to both the limited Project Team user interviews and the two surveys conducted by SWIRS Management, users are generally satisfied with both quality and turnaround (although service is less timely than demanded by SWIRS). Only three queries of 500 were returned without search results.

Several factors should be examined in any effort to improve performance:

- Selection of Response Materials

  As mentioned in MANAGEMENT PLANS AND OBJECTIVES there is a lack of selectivity by the contractor in determining the level of response warranted by the subject, format, and type of request. For example, a High School student requested information for use in a school debate. The request identified nine different subject areas and generated four canned searches, five computer searches, 27 xeroxed pages, and an additional reference list. While the letter was somewhat unusual in its breadth of subject coverage and purpose, it is representative in the fact that no judgement of request source was made. A number of similar instances were observed. In general, the number of documents mailed to satisfy a general request might easily overwhelm a first-time user.

- Number of Citations per Request

  The analysis of sample searches revealed a number of instances in which very large numbers of citations were selected. It was not unusual to discover several hundred 'hits' for a single search. Since the screening activity has been discontinued for non-EPA users, this characteristic might have a serious effect on the service rendered for any new (not canned) search. It may also be an indication that the SWIRS thesaurus, containing 700 terms, is too broad.

- Quality of Hard Copy

  While the general public currently receives computer printouts of abstracts (since they are no longer screened), EPA users are still screened and therefore receive abstracts on microfilm printer copies. Computer-printed copies are of much better quality than microfilm copies, which several interviewers considered poor. This is really a conflict of screening versus paper quality, since it would be more expensive to screen and then generate computer copy of selected abstracts (of course, an on-line system would provide this benefit). Users were happy with the design of the abstracts and the fact that there was one abstract per 8 x 11 inch page, whether computer or microfilm generated.

- Multiple Responses

  In a large number of instances the Project Team observed that the EPA Hq Form 8700-4 did not include a statement to the effect that additional material will be forthcoming. In the likely event that a request generated more than one computer search, it would seem to the user that a portion of the request was not accomplished. A form note to this effect should be added to the EPA Form.

- Canned Searches

  A canned search is a group of preassembled materials. For example, a query on the production of Methane gas would not require a computer search. Instead, the preassembled material would be mailed. However, the Project Team observed that the canned search material for "Solid Waste Management, General" contains 57 xeroxed pages of bibliographic abstracts. According to the contractor, most canned searches contain more. Arthur Young and Company believes that general requests might better be served through a few pamphlets on the subject such as those produced by the OSWMP Technical Information Staff's Publishing Operations Branch and those other EPA and other government publications which are provided without charge. Those users with a requirement for more information can then formulate and submit more detailed queries.

- New User's Packet

  Information of a general nature explaining the service is sent along with the output to first-time users. This packet contains several booklets explaining the functions of the retrieval service, other library holdings, a general reference list, a user's guide, and an information folder describing the service provided by the Air Pollution Technical Information Center (APTIC). Upon request a thesaurus of keyword index terms will be included. Since the intention of the new User's Packet is to inform first-time users of the services available, Arthur Young and Company believes that a simple, two or three page brochure would be more suitable.

  The brochure would briefly explain the purpose of SWIRS, but would primarily be used to delineate the subject matter covered and provide instructions designed to enable the user to formulate more detailed queries, complete with the purpose for which the information is required. Users would still be afforded the opportunity of ordering the thesaurus.

The Project Team did note that the APTIC brochure being sent was frequently a xerox copy of relatively poor quality, and that the brochure was apparently outdated. The printed version of the brochure, with its attached reply form, would be more useful

- Additional EPA Source Material

As mentioned in the MAJOR RECOMMENDATIONS AND ACTION PLAN, an important source of EPA material is not being provided to interested users of SWIRS. The Office of Research and Development produces a number of manuals designed as "how-to' guides for industry, municipal government and others under its Technology Transfer program. Some examples of these manuals include:

- Municipal Seminar Publications
  - Physical - Chemical treatment
  - Oxygen Activated Sludge
  - Nitrification/Denitrification
  - Upgrading Existing Wastewater Treatment Facilities - Case Histories
  - Wastewater Filtration
  - Air Pollution Aspects of Sludge Incineration
- Industrial Seminar Publications
  - Upgrading Poultry Processing facilities to reduce pollution
  - Upgrading Metal Finishing facilities to reduce pollution
  - Upgrading Meat Packing facilities to reduce pollution
  - Upgrading Textile Operations facilities to reduce pollution

These manuals are available free of charge, and would appear to be an excellent reference source for requests regarding covered subjects.

(4) Data Processing Characteristics

The SWIRS data processing system was designed in 1970 and, by today's standards is relatively inefficient. Under normal conditions some limited improvement could be made to the system design, to make the system slightly less inefficient. However, since all source documentation and program source libraries were lost during the time they were stored at the National Institutes of Health, it is impossible to modify existing programs.

The primary areas of system deficiency include:

- Sequential File Design - In earlier versions of Operating System Design, there existed substantial cost difference between Sequential and Direct, or Random Accessing, with Sequential Accesses being substantially cheaper than a corresponding number of Direct Accesses. Consequently, economics frequently forced the utilization of sequential processing, even though this method required the program to read through all preceeding records to reach the one desired. Since that time, the cost of Direct Accessing has decreased to a level that is only slightly higher than Sequential Accessing. The result is that a sequential system such as SWIRS, which retrieves a limited number of records from a fairly large file, is substantially more costly than a similar Direct Access System. While in some limited number of instances it is possible to economically modify a sequential system to employ Direct Accessing, the fact that all source material has been lost precludes this approach with SWIRS.

- Concurrent Searching - The SWIRS system, run in batch mode, was designed to allow several searches to be run against the data base at a single time. However, this capability appears to be limited to a maximum of 6 searches, since program design requires each search strategy to appear on a separate input file, and accepts a maximum of 6 input files (search outputs each appear on a separate file also). Moreover, the execution of more than even 3 or 4 searches concurrently frequently causes program blowups, presumably because the internal table on which the program places search strategies is of limited size. Of course, this problem would become more costly as use of SWIRS increases. Without source documentation and libraries, the search program cannot be modified, although a new program could be written.

- Batch Processing - The data processing system provides no online search, inquiry or retrieval capability. The ability to query the data base interactively is an extremely valuable tool, particularly when the search specialist is not confident that his strategy will produce the desired results. While interactive processing is considerably more expensive than batch processing, it should at least be available for special situations. At the present time, the program operator must run all batch searches to a WYLBUR File before printing. This file is then queried through WYLBUR to determine its size. If too large, the file is not printed, but rather the searcher is notified to enable him to refine his strategy and make a second run. Since most data processing is performed at odd hours because of system traffic (frequently on Saturday), this problem can considerably delay response to the user, particularly if several search-runs are made. Finally, the searcher has no opportunity to verify the relevance of material retrieved until the batch run is completed. It is probable that the ability to interactively query the data base would reduce or eliminate the need to manually screen search results. The addition of an interactive search capability would require a total redesign of the data processing system. The increased cost of an online capability must be weighed against the benefits of more accurate and timely retrievals and reduced personnel screening activity.

- System Flow - The design of the system requires an excessive amount of manual intervention to load files, examine outputs and execute individual programs. This appears to be due to a number of factors including obsolete system design, lack of operator confidence and the non-existence of a proper job control stream. This excessive operator intervention is the prime factor causing the system to be run at odd hours, since it is primarily these activities that could be affected by a high amount of data traffic and computer utilization. While the development of an optimum job stream might be precluded by the lack of source libraries, some improvement could be made after a review of the various job control functions and file preparation activities. It does appear, on the surface, that more effort could have been taken during system conversion to OSI to create a more efficient job control system, rather than merely converting the existing commands. If the current system is retained, the system job stream should be examined and modified to improve ease of operation and reduce manual intervention.

- Error Diagnosis - There appears to be only moderate understanding of the system design by all persons concerned with its operations. This could be due partly to a lack of research by operating personnel. The Project Team observed a program blowup for which the operator had no explanation and was forced to leave for the searcher to determine, presumably by re-evaluating his search strategy. While better research of the system would increase understanding, the lack of proper diagnostics from the program and the lack of any edit of search strategies will continue to cause difficulty.

- WYLBUR Expense - The use of WYLBUR as an input tool is an extremely expensive technique. This problem, and alternative approaches, is discussed in Appendix G. It should be noted that there is nothing in the SWIRS design that would preclude the use of keypunch, keytape or typewriter input (MC/ST) devices, since a simple program could be written to modify such input to be identical to that required by program design.

(5) Source Overlap

The ANALYSIS OF REQUIRED JOURNALS, Appendix C, describes the analysis performed on the required journals of each of the Technical Literature Systems. However, it should be noted here that a substantial overlap exists between those journals whose screening is required by SWIRS, those which are required by APTIC and Pesticides, and those which are normally screened by the various commercial abstracting services. The following table illustrates the percentage of the SWIRS required journals also used as sources by the other services reviewed:

SERVICE	PERCENTAGE OF SWIRS JOURNALS REVIEWED
Chemical abstracts (CA)	62.1%
Engineering Index (EI)	24.8%
CA or EI	71.2%
APTIC	36.3%
Pesticides	15.1%

The fact that a required SWIRS journal is reviewed by another service does not prove that the same individual articles are selected. However, since there is such a great degree of overlap, there is a definite possibility of a considerable degree of overlap in selected documents, particularly since Chemical Abstracts and Engineering Index do include SWIRS-related subjects in their selection criteria. In addition, the high degree of source overlap among the required journals of the EPA systems suggests that a combined screening and acquisition activity would be beneficial.

The Project Team selected 10 actual searches at random for comparison of data contained on the SWIRS data base with that available commercially. Time and cost considerations forced the Cincinnati Library, the organization performing the searches, to perform only two of the selected searches. The results of these two searches included very few duplicate articles. The commercial system provided a sizable number of citations on the more general of the two searches, but very few hits on the more selective search. As discussed in the RECOMMENDATIONS AND ACTION PLAN Section, these results dictate that caution be used in any assumption of the existence of parallel or duplicate contents of data bases.

SEARCH SUBJECT	NUMBER OF CITATIONS SELECTED					
	SWIRS	NTIS	CA	EL	POLL.	DUPLICATES
Baltimore Monsonto Landgard Process (after 1974)	14	1	0	2	1	1
Site Selection for Regional Solid Waste Operation	30	78	59	32	N/A	2

4. COMMERCIAL ABSTRACTING SERVICES

Substantial cost savings could be realized through the use of commercially developed data bases supported by commercial data processing systems. The following analysis portrays the costs which would be experienced through such an approach, and anticipates a reduction of annual SWIRS costs of $ 65,774, due to improved data processing capabilities.

The cost columns depict anticipated costs for:

- No usage of commercial systems, but substantial improvement to the SWIRS data processing system.

- Total reliance on commercial systems, but with continued selection and acquisition activity equal to half of the current effort. The resulting data base would be donated (or sold) to one of the commercial abstracting services.

- Total reliance on commercial systems. Note: Chemical Abstracts, for example, has expressed interest, as a condition of receiving the current SWIRS data base, in expanding its own selection activity to include the SWIRS criteria.

		* * COMMERCIAL SEARCHING * *	
COST COMPONENT	NO COMMERCIAL USAGE	HALF CURRENT ACCESSIONS	NO ACCESSIONS
SWIRS COST			
Contractor Search Charge($35/search)	42,000	-	-
Data Processing($10/search)	12,972	-	-
Data Base Creation($239,882)	239,882	119,941	-
Data Processing Input($38,923)	38,923	19,462	-
TOTAL SWIRS	333,777	139,403	-
COMMERCIAL SERVICE			
TOTAL CHARGE($50/search)	-	60,000	60,000
TOTAL COST	$333,777	$199,403	$60,000

If the number of searches performed by SWIRS should triple to 3600 annually, and if SWIRS computer cost/search were to decrease from $10 to $5 due to economies of size, the following costs would result:

COST COMPONENT	NO COMMERCIAL USAGE	* * COMMERCIAL SEARCHING * *	
		HALF CURRENT ACCESSIONS	NO ACCESSIONS
SWIRS COST			
Contractor Search Charge($35/search)	126,000	–	–
Data Processing($5/search)	18,000	–	–
Data Base Creation($239,882)	239,882	119,941	–
Data Processing Input($38,928)	38,923	19,462	–
TOTAL SWIRS	442,805	139,403	-
COMMERCIAL SERVICE			
TOTAL CHARGE($50/search)	–	180,000	180,000
TOTAL COST	$422,805	$319,403	$180,000

The figures used in the above analysis are the most conservative which can be developed from the changing and sometimes conflicting data acquired. As such they considerably understate the probable monetary savings which would result from these alternatives.

The following factors have impacted the development of the figures used in the above analysis:

- The Franklin Institute Cost Proposal for April-June, 1975 contained an allocation of $55.97 per search for the manual activity necessary to prepare a search from a user request. The corresponding allocation for October-December, 1975 was only $35.36 per search. Our analysis assumes a cost of $35 per search. It should be noted that these costs per search are allocated by Franklin Institute, the contractor does not base his charges on any type of piece rate. Finally, the total cost of this later proposal was considerably greater than for previous quarters due to proposed increases in four of the other five activities, (SWIRS Library Maintenance, an overhead activity, being the only one which did not increase).

- There exists no breakdown of NIH computer charges between monies spent for computer searching versus update and maintenance activity. Recognizing the high cost of WYLBUR entry as used to enter new abstracts into this system, we have allocated only 1/3 of data processing charges to computer searching resulting in a per search charge of $32.44. Further, since the continued use of the SWIRS system should include the design of a new, efficient data processing system, we have allocated only 1/3 of that cost for this analysis, resulting in a cost of $10.81 per search. The analysis also assumes that the new data processing system would reduce data input and maintenance costs by 50%.

- The Franklin Institute Cost Proposal for April-June, 1975 assumes a total of 1440 searches per year. SWIRS operations counts for the first 9 months of 1975 provides an extrapolated value of 919 annual searches. The Project Team analysis provides an extrapolated figure of exactly 1200 searches annually. This analysis assumes a value of 1200 annual searches which is considered by the Project Team to be the most realistic figure available.

- The cost allocated to a search of commercial data bases is taken from independent research performed by Mrs. Mary Calkins, Information Retrieval Librarian at the EPA Cincinnati Library. The methodology and results of this research is currently in preparation for publication. This research, performed using actual searches performed by the Cincinnati Library, resulted in an estimated cost of $15.68 per search on any single commercial system (free government-operated systems were not included in the study). During the period examined, an average 2.7 data bases were queried through the single data processing system, using an average of 20 minutes connect time. Additional costs, including terminal costs, telephone costs and one hour of search specialist effort add $10.63 for a total search cost of $26.31. The Project Team has nearly doubled this figure (to $50/search) in order to ensure a conservative comparison.

5. PRIMARY CONSIDERATIONS

It would seem that there would be a strong National interest in a Solid Waste Management Information Retrieval System. However, actual usage figures do not support this assumption. According to the Project Team Sample Search Analysis, a total of 1200 searches were made during the latest one-year period, generated by only 624 requests for information. Consideration must be given to the justification of the expenditure of more than $500,000 in satisfying 624 requests. In addition, few of these 1200 searches represent the population which SWIRS is most interested in assisting. Private Citizens, Citizen Action Groups and Students represent only 293, or 24% of the past year's searches.

OSWMP is faced with a rather complex dilemma. A great deal of effort has gone into the creation of an information source which apparently can effectively respond to the vast majority of queries on the subject of Solid Waste Management, but this source is not only very expensive, it is not being utilized. If the Office of Solid Waste Management Programs is to fulfill the requirements of the Solid Waste and Disposal Act as it has interpreted them, it must take steps, and incur expense, which will increase the level of recognition and usage of SWIRS services. Since OSWMP cannot increase its expenditures for SWIRS (and in fact desires to reduce them), the funds for any increase in publicity and training must come from a reduction in other SWIRS activities. While the development of a more efficient data processing system and the elimination of WYLBUR would reduce total costs, it would be unrealistic to anticipate an annual savings greater than $50,000 from improvement in system design.

III. NOISE INFORMATION SERVICE

1. SYSTEM DESCRIPTION

The Noise Information Service (NOISE) is a computerized bibliographic system containing approximately 7,000 document abstracts and citations. The system was developed by the Office of Noise Abatement and Control (ONAC) in response to the Noise Control Act of 1972 which set forth a policy of "effectively coordinating federal research and activities in noise control; establishing federal emission standards for products distributed in commerce; and providing information to the public respecting the noise emission and noise reduction characteristics of such products." This Act also requires other agencies to consult the EPA before prescribing noise regulations and allows the EPA to publicly review any regulations which it considers potentially detrimental to public health and welfare.

An "Analysis of Information Requirements of the Noise Abatement Community" was performed in the spring of 1972 by Informatics, Inc. under contract with ONAC. The report concluded that the actual or potential community of information users was heterogeneous and required a variety of services to satisfy their needs and that by far the largest user group were individuals having "a continuing, but partial, interest need, or who have a one-time, partial interest need." As a result of this study and ONAC's desire to serve this potential community users, ONAC decided to develop the NOISE bibliographic system. Within this user community, ONAC expected a large percentage of system usage to come from EPA contractors, from EPA staff preparing congressional reports, and from the Agency's assistance programs for State and local organizations.

The day-to-day responsibilities for operating the system are presently contracted totally to Informatics, Inc. Their services include all of the normal functions performed by input contractors (e.g., acquisitioning, screening, abstracting, indexing, and input preparation) as well as the responsibility for managing the data processing operations and the system's document collection. Informatics screens the world's literature (approximately 200 journals) and inputs only the best of the world's literature which pertains directly to noise. The selection criteria allows only information considered by Informatics to be of long-term value to enter the system. As a result, much of the information required by ONAC to set standards, such as State and local ordinances, frequently is not available through NOISE. To obtain this kind of information, ONAC expanded its information resources in June, 1975, by developing its Shared Information System (SIS). All of the data which currently flows into NOISE is also being input to the SIS, as well as any information ONAC feels is necessary to satisfy its regulatory responsibilities. This additional information includes such data as product standards, industry profiles, correspondence, and legislative and regulatory statements, and is selected by ONAC itself from information acquired during normal operations. The SIS is actually both a bibliographic system and an automated filing system. Although together both systems provide an information retrieval system, there are actually two separate information flows which result in two separate data bases, each accessed in a different manner. The SIS presently involves processing an automated master file from which a Shared Acquisitions List is periodically printed. All searching and selection is performed by manually reviewing this document using various indexes. The NOISE System, on the other hand, provides users with a sophisticated on-line, interactive search capability of the NOISE file. Informatics, Inc. is the present contractor for both NOISE and SIS.

The funding responsibility for both Informatics contracts has recently been transferred to the ONAC Standards and Regulations Division, which originally developed the Shared Information System. With this transfer will come a fundamental change in the characteristics of the NOISE service. This system shall be directed more toward ONAC operational requirements and less toward potential outside users, and will combine the functions currently performed by NOISE and SIS. Rather than providing detailed abstracts of documents judged to be of lasting value, the system shall provide very short abstracts (with quite detailed indexing) of documents considered necessary for current ONAC operations. In addition, selection of included documents will be not through contract or selection, acquisitioning and abstracting service, but through the normal activities of the EPA. Such information would include:

- Finished documents generated during contractor projects, indexed in more detail than currently provided by NTIS.
- Pertinent contractor working documents and external information gathered by contractors during the performance of a project.
- Documents submitted by private industry during legislative or regulatory hearings.
- Information gathered or generated, and adjudged pertinent, by ONAC staff.

The current NOISE data processing system utilizes a modified version of the STIMS/RECON software system developed by NASA for on-line computer searching of random storage files. This sophisticated system is fairly representative of the current state-of-the art in bibliographic technology. The system provides on-line access to its thesaurus as an aid to searchers and allows access to the master file by author, corporate source, subject term, and date of entry through four inverted files. English-format commands are used by terminal operators to access the data base in a conversational manner. That is, once the requestor has gained access to the system, the computer requests commands, and the user responds with appropriate command words. The system responds in turn with requested information and concludes with a request for another command. Users unfamiliar with the system's thesaurus can query the data base and by using the "Expand" command, determine related subject terms, more general terms, and more specific terms. A search is performed by specifying Boolean combinations of subject terms for which the system replies with the number elements in the data base matching the user's strategy. The user can then further refine his strategy or print all or part of his selections either through the terminal itself or through off-line batch printing. The latter option may be more desirable if the output volume is large due to the high costs of on-line activity.

The system will also allow the user to scan titles only, if he desires to look at all of the search results while on-line but does not need the entire abstract or citation. STIMS/RECON also has the option to perform "free-text" searching. That is, it can search on any word whether it is in the system thesaurus or not. This option, however, is not utilized in the present NOISE System because it increases the size of inverted files as well as requiring additional search time which together can greatly increase the system's costs.

2. MANAGEMENT PLANS AND OBJECTIVES

Responsibility for the NOISE system has recently been transferred from the Program Management staff to the Standards and Regulations Division, whose management's evaluation of the system is consiserably different from that of previous management. While previous management felt that the system should provide technical assistance to the regions and the outside world, present management believes that the system should primarily support the regulatory functions of ONAC. It considers the current data base to be too limited and obsolete to effectively provide this support.

The information requirements and program goals of present management are discussed in detail in the PROGRAM MANAGEMENT OBJECTIVES REPORT produced under this Task Order. The primary Objectives are summarized below.

(1) Management believes that the present costs of the NOISE contract are excessive in relation to the services provided, and is seeking means of reducing those costs. Several steps are being taken to effect immediate reduction, including:

- Elimination of Special Tasks performed by the contractor for the NOISE System for particular studies performed by previous System Management.
- Reduction in length of abstracts prepared by the contractor.
- Elimination of subjects which would normally be included in the SIS.

(2) Management desires a single bibliographic system which would perform the current functions of both NOISE and SIS, and is planning to issue a Request for Proposals to analyze the information needs of ONAC and determine the type of system most useful in fulfilling these needs. It is not anticipated that a totally new system will be necessary; rather management expects to either purchase a packaged system or utilize a data base management system. Arthur Young and Company believes that such a system should be designed to also support the data processing requirements of the other EPA Technical Literature Systems. While the NOISE/SIS data base may be somewhat different in its contents from traditional bibliographic files, the system could be developed to support the needs of all the systems, if cooperative effort is quickly initiated.

(3) Management believes that almost all information pertinent to the needs of ONAC is either generated or noted by ONAC long before it enters the world's literature. This is a primary rationale for the planned reliance upon ONAC staff and contractors noted in the System Description.

3. ANALYSIS OF OBSERVED CHARACTERISTICS

This discussion addresses those characteristics of the NOISE service which the project team considers to be most germane to the purposes of this study. Throughout the conduct of the study, the Project Team was unable to identify an individual who might be considered the NOISE System Manager for the purposes of this study, not only because the system is so completely run by contract personnel, but also because it is both changing hands and in the process of being eliminated, at least in its present form. In addition, the NOISE service is by far the most anonymous of the EPA Technical Literature Search Systems, since even its users frequently do not know whence their information came. Consequently, the Project Team was not able to research the NOISE Service to the same depth as the other systems.

The Shared Information Service was not included in the scope of work to be performed during this project, and time did not permit a detailed analysis of that system. However, where particular characteristics of the SIS were noted by the Project Team, they are referred to in this discussion.

(1) System Usage

From October, 1974, through September 1975, there were 460 NOISE terminal sessions, as recorded on the RECON Statistics Report generated from the system for Informatics. A terminal session is defined as the period of time when a user "logs on" his terminal to access the data base until he "logs off." During this time more than one search may be run but usually each session corresponds to a user's attempt to acquire information on a specific topic and is therefore roughly equivalent to a search request on some of the other systems.

Since all searching is performed from on-line terminals, it is relatively easy for the system to collect usage statistics by user (if this capacity is included in system design). Each user has a unique ID number which precludes unauthorized users from accessing the system, and which allows the system to automatically accumulate by user the number of sessions and amount of time spent in accessing data. The following table provides a breakdown of these statistics for each user accessing the system from October, 1974, through September, 1975:

	Terminal Sessions		Duration	
	No.	%	Min.	%
Informatics	60	13.0	710.34	12.0
ONAC Headquarters*	323	70.0	3,881.55	70.0
Region 1*	3	0.7	44.15	0.8
Region 2*	11	2.4	325.26	5.9
Region 5*	6	1.3	48.30	0.9
Region 9*	6	1.3	323.61	5.8
Region 10*	1	0.2	49.81	0.9
Cincinnati NERC*	17	3.7	27.13	0.5
Projects-Washington*	1	0.2	7.46	0.1
Air Pollution Control Washington*	1	0.2	0.15	0.0
Environment Canada	3	0.7	1.42	0.0
Weire & Company	1	0.2	0.17	0.0
Collopan Corporation	22	4.8	67.81	1.2
Bellomy	5	1.1	53.09	1.0

---

* EPA Users.

Perhaps a more meaningful measure of this usage would be to relate it to every day terms. This total activity represents 1.26 requests/day for an average of 12 minutes/request or approximately $234/request ($107,803 NOISE costs associated with NOISE divided by total searches). The above figures indicate that EPA directly accounts for 80.2% of total usage although the 13% attributed to Informatics is for activities in support of EPA. Five percent of the total can be attributed directly to industry (Weire & Company and Colopan Corporation), and less than 2% to academia (University of New Mexico) and other governments (Environment Canada). Bellomy is a West Coast account that demonstrates the NOISE System to potential subscribers as part of Informatic's marketing efforts and accounts for only 1.1% of total usage. Documentation is not available concerning end users who may have initiated requests researched by these subscribers. However, both present and previous management feels that the system principally serves EPA (i.e. EPA personnel are end-users) and recognizes that there is very little awareness of the system outside of the agency.

In May, 1974, the regions were requested to discontinue accessing the system directly and to channel their requests through ONAC Headquarters. Therefore the only statistics relative to the regions' usage of the system are for the period October, 1974, through May, 1975. The regional usage (15.43%) is significant because it represents an approximation of the usage which may be attributable to non-EPA requestors. Although the regions are sometimes end-users themselves, much of their function is to provide assistance to State and local organizations. Within the Washington D.C. Office of Noise Abatement and Control, the Technical Assistance and Operations Division also queries the NOISE System in assisting any outside user who requests information. However, since everyone within ONAC Headquarters uses the same User ID, there are no statistics which show what portion of the total ONAC usage is attributable to that division and therefore to end-users outside EPA. Most employees contacted during the user survey stated that they did not use the system more than once or twice a month.

(2) Operations Costs

The EPA TECHNICAL LITERATURE SEARCH SYSTEM PROFILES contains a breakdown of NOISE costs, provided by the contractor, which depicts expenditures for the year ending October, 1975 by the two contracts under which Informatics services ONAC:

Contract	Cost Component	Cost	
6801-3114	NOISE OPERATIONS	46,467	
6801-3114	SOUNDINGS Publication	13,582	60,049
6801-3115	Special Tasks	111,473	
6801-3115	SIS Library	67,563	
6801-3115	SIS Publishing	2,372	181,408
			$241,457

Exhibit III figures for Fiscal Year 1975, ending June 30, retrieved from the 1975 SUMMARY OF CONTRACT COSTS of the Financial Management Division, depict quite different costs for both contracts, with contract 6801-3114 (NOISE Operations and SOUNDINGS) costing $121,385 and Contract 6801-3115 (Special Tasks and SIS) costing $128,458. The cost for NOISE Operations alone can be used to develop the cost per search of current NOISE activity.

COST COMPONENT	CHARGE
Contract 6801-3114	$121,385
- SOUNDINGS Publication	- 13,582
NOISE operations	$107,803
÷ Number of Terminal Sessions	÷ 460
Cost/Terminal Sessions	$234.35

NOTE: SOUNDINGS Publication costs are removed from the calculation since they are not required for the operation of NOISE.

(3) System Service

Based upon the user interviews described in Appendix B, the project users who are familiar with NOISE consider the information in the data base to be of good quality and serving a useful purpose. However, the overall system was considered to be of limited value because it satisfies only a very narrow segment of user's information needs. Management recognizes this problem and plans to alter the scope of literature covered. Conceptually, management anticipates changing the content of material going into the system from journal acquisitions to contractor reports and the results of their literature searches, materials selected by ONAC staff which they feel is pertinent to their work, and papers and reports supporting regulatory and judicial activities. The current review of literature performed by the contractor will be discontinued. It is hoped this methodology will provide a greater amount of detailed information on those subjects in which ONAC is most interested. Since the current scope of literature coverage does not appreciably support ONAC or its contractors (who use the present system very little), management does not feel that the elimination of the contractor selection activity will significantly impair the ability of ONAC to discharge its responsibilities, even though the selections are considered to be of high quality.

Several particular points can be developed from the results of the User Interviews (Appendix B) and sample Search Analysis (Appendix C):

- Limited Awareness of System

  Approximately one third of those users contacted were not familiar enough with the NOISE system to answer questions regarding its performance.

  To most of this group the services provided by NOISE and the Shared Information System (SIS) are indistinguishable because of the growing practice of channeling all requests for literature through the ONAC Library run by Informatics Inc. The libriarian, who is familiar with both systems, usually determines which system can satisfy a user's request. Providing a single source of information to satisfy all requests is a suitable procedure, particularly since the professional staff is also allowed access to the computer terminal whenever their needs are more general. However, the fact that several of the staff who were contacted during a limited user survey were not even aware of a computer search capability reflects upon a serious lack of visibility within ONAC concerning the NOISE service, and demonstrates a need for better, more formal orientations for employees.

- Turnaround

  All interviewers who were aware of the system were pleased with its turnaround. In most cases turnaround for large volume printing requires no more than two or three days but may be obtained sooner if there is a need. Of course, there is only a delay if the output volume is large enough to warrant printing the abstracts or citations off-line. If the search results are small in number, they can be output on the terminal, and turnaround is immediate. For these reasons, NOISE provides much better turnaround than any other EPA bibliographic system.

- Limitation of Scope

  All interviewees reported that they utilized the system when it served their purpose, but did not use it often because the kind of data available in the system was not frequently needed. Examples of information desired and not present on the data base include descriptions of particular ordinances and older published material.

- Quality of Abstracts

  Users frequently were pleased with the length and detail of the NOISE abstracts. In fact, some users stated that they could sometimes gather all of the information they required from the abstract alone. The problem associated with this characteristic is that, as current Program Management has stressed, the usage of such limited information without in-depth review of the entire article can result in conclusions tangential to, or in conflict with, the overall document, or unsupported by the research referenced by the document.

(4) Data Processing Characteristics

Because the NOISE data processing system is a proprietary system owned by Informatics, Inc., the Project Team did not perform a detailed analysis of the automated system, nor were internal operations problems, if any, discussed with the contractor. The review of system capabilities and demonstration of system performance has convinced the Project Team that this is by far the most sophisticated and inclusive of the data processing systems reviewed, and should be capable of supporting the primary requirements of most traditionally designed bibliographic systems. No attempt was made to evaluate the efficiency or cost of the data processing system.

(5) Source Overlap

The ANALYSIS OF REQUIRED JOURNALS, Appendix C, describes the analysis performed on the required journals of each of the Technical Literature Systems. A substantial overlap exists between journals reviewed by Informatics, those reviewed by the other Technical Literature Systems, and those reviewed by several of the commercial abstracting services. The following table illustrates the percentage of NOISE journals also reviewed by the other services:

Service	Percentage of NOISE Journals Reviewed
Chemical Abstracts (CA)	52.7%
Engineering Index (EI)	19.9%
Biological Abstracts (BA)	32.2%
CA or EI or BA	66.4%
APTIC	24.7%
Pesticides	13.4%
SWIRS	12.6%

The fact that a required NOISE Journal is reviewed by another service does not prove that the same individual articles will be selected. However, since there is a considerable degree of overlap, this possibility must be recognized.

Two actual NOISE searches were selected and run on several of the commercial data bases to provide a further indication of the potential for documents overlap. While a reasonable number of citations were returned by these data runs, only a few citations duplicated those returned by the NOISE system.

SEARCH SUBJECT	NUMBER OF CITATIONS SELECTED						
	NOISE	NTIS	CAIN	BA	POLL	MEDLARS	DUPLICATES
Vegetation Attenuation of NOISE	43	0	19	5	11	N/A	10
	Note: 6 of the 10 NOISE articles duplicated appeared on more than one commercial data base						
Physiological effects of quarry mining and construction noise	25	12	N/A	N/A	N/A	100	2

4. PRIMARY CONSIDERATIONS

ONAC Management is planning to perform a detailed analysis of its information needs, and final determinations of system design and contents have not been made. Arthur Young and Company concurs with Management's assessment that service costs appear excessive, and with the decision to re-evaluate program and system requirements. Since the new system will act as much as an Automated Filing System as a Technical Literature System, Arthur Young and Company is unable to adjudge its necessity. However, the limitation of selected documents to those actually used by ONAC will avoid the collection of a large amount of extraneous information.

In addition to the managerial and operational factors inherent in such a development effort, the following considerations should be addressed:

(1) Information Dissemination Policy

In the RECOMMENDATIONS AND ACTION PLAN of this report, the need for an Agency-wide policy regarding information dissemination is identified and discussed. Since the new NOISE System will be oriented towards the operational requirements of ONAC, care should be taken to ensure the capability to identify and respond to external user needs, should EPA policy so require.

(2) Usage by Other Systems

Since ONAC is planning to develop a fairly comprehensive system, it would be of value to all of EPA if consideration is given during design to the characteristics of the other systems, and to the generalized requirements of a complete bibliographic system. This effort, performed in coordination with the Management Information and Data Systems Division, might provide a product which could, at slightly higher development cost, be used to process any of the EPA Technical Literature Systems or other EPA bibliographic systems.

IV. AIR POLLUTION TECHNICAL INFORMATION CENTER

1. SYSTEM DESCRIPTION

The Air Pollution Technical Information Center (APTIC) has responsibility for the collection and dissemination of literature on air pollution to as wide an audience as possible in order to fulfill the requirements of the Clean Air Act of 1963. The act specifically authorizes a central agency to "collect and disseminate in cooperation with other Federal departments and agencies, and with other public or private agencies, institutions, and organizations having related responsibilities, basic data on chemical, physical, and biological effects of varying air quality and other information pertaining to air pollution and the prevention and control thereof." In fulfilling this role, APTIC provides free literature searches of its computer files to any requestors and encourages expanded usage of the system. Although the published literature in the field of air pollution is extremely prolific, APTIC attempts to sample as large a cross section of published journals as possible. Currently over 7,000 sources of information are regularly screened, 1,100 of which are considered high yield sources and must be included in this screening. As a result, APTIC is considered by management to be a clearinghouse of information relating to air pollution.

Since the Office of Air Quality Planning and Control does have regulatory responsibility, much of EPA's usage stems from the regulatory programs, standard setting efforts, strategy planning, and research and development endeavors. APTIC is also a resource for State and local organizations as well as the regional EPA offices. Program management is particularly concerned that APTIC provide information to those people who have no alternative resources for acquiring it.

APTIC is presently located at Research Triangle Park, N.C., and utilizes the Univac 1110 computer. APTIC personnel have recently been reduced to one person, the System Manager, who has Primary responsibility for the day-to-day management of the system as well as formulating search strategies for users. The System Manager's former assistant is currently with APTIC awaiting reassignment. Assistance in such tasks as xeroxing hard copies is provided by a clerical pool shared with other services. A contractor, D.P. Associates, is presently responsible for interfacing with computer operations and maintaining the software system. The input contractor, Franklin Institute in Philadelphia, screens the world's literature and provides bibliographic citations and abstracts in machine readable form as well as hardcopy and microfilm copies of accessioned documents. APTIC maintains a document collection from which it provides free copies of any documents requested by EPA or its contractors. During 1975, 4,700 documents were provided to requestors. In addition, APTIC periodically publishes selected bibliographies, abstract bulletins, and catalogs which are mailed to over a thousand recipients and are available to any other requestors. The input contract is to be let for rebid in the near future, and because of the personnel reductions imposed upon APTIC, the responsibilities of the contractor will be expanded to include interfacing with users and submitting search strategies.

The data processing system is a batch application which was developed and implemented by EPA's Data Systems Division specifically for APTIC on the Univac 1110. Although bibliographic citations are available on-line to regional users, abstracts can only be retrieved by submitting a batch search. This can be accomplished by writing or telephoning APTIC and having someone resubmit the search strategy, or document accession numbers if they are available. The data base includes an inverted file which allows access to the master file via Boolean combinations of keyword terms, a master file comprising eight reels of tape (over 70,000 records), and various other secondary files. A user may access the system via subject terms from a hierarchical thesaurus. The system allows for additional searching capability through an open-ended card file which can be manually searched for approximately 3,000 terms which are not included in the system's thesaurus. Each open-ended term has the accession numbers recorded which correspond to documents in the data base referencing that term. Additional information describing the system may be obtained by referencing the EPA TECHNICAL LITERATURE SEARCH SYSTEM PROFILES, produced under this Task Order.

2. MANAGEMENT PLANS AND OBJECTIVES

Program Management plans, needs, and objectives for the Office of Air Quality Planning and Standards are documented in the "Program Management Objectives Report" developed under this Task Order. The primary considerations of management are summarized below:

(1) Expansion of System Usage

Although APTIC is believed to be well known to State and local environmental agencies and to the ecological community as a whole, management believes that emphasis should be given to increasing that recognition and expanding system usage. While little effort has been made to increase the system's visibility, management feels that some marketing effort would be both desirable and appropriate. Arthur Young & Company concurs in this assessment, noting that all the EPA participants in the project's user survey had learned of APTIC through informal communications with other EPA employees. Several non-EPA users surveyed had learned of the service from friends or business associates rather than through the direct efforts of EPA to expand potential usage.

(2) Limited Visibility From Program Management

Management expressed satisfaction with the current level of service, but acknowledged that the service functioned without much attention from program management. Although confidence in the system manager was expressed, management also recognizes the need for a formal check upon the quality of performance. This point bears emphasis since it would be fairly easy for the quality of the service to decline without management recognizing it. In fact, this appears to have already occurred, since system usage has been drastically declining and data processing problems have degraded service response time, as identified in the ANALYSIS OF OBSERVED CHARACTERISTICS. This lack of detailed performance review is common to several of the systems and is discussed in the RECOMMENDATIONS AND ACTION PLAN of this report.

(3) Personnel Reductions

Management has reduced the staff at APTIC to one in order to satisfy personnel requirements in other areas of OAQPS. The functions performed by the displaced personnel are to be performed by the input contractor under the terms of the forthcoming contract. Although management does not intend to de-emphasize the functions performed by APTIC, this decision nevertheless restricts the control which it is able to exercise over the system.

3. ANALYSIS OF OBSERVED CHARACTERISTICS

This section of the report identifies those system characteristics, both strong and weak, which the Project Team feels are pertinent to the evaluation of the APTIC System. The purpose here is not to provide a description of the system, which may be found in SYSTEM PROFILES, but to identify those areas where specific improvement can be made and to provide the basis for an overall review of the system. The information presented in this section of the report will serve as a basis for the recommendations provided in the RECOMMENDATIONS AND ACTION PLAN.

(1) System Usage

APTIC presently supports approximately 876 annual requests for literature searching which result in approximately 1300 computer searches. These figures are extrapolated from the number of search requests (803) tabulated by APTIC from January, 1975, through November, 1975, and from statistics compiled during our analysis of user requests (Appendix A), which revealed an average of 1.5 computer searches per request. According to these figures and to figures supplied by the system manager, the volume of requests has continued on a downward trend over the last three years.

Year	Requests
1972	1941
1973	1680
1974	1277
1975	876

A three month sample of literature search requests for 1975 was drawn to determine the identity of the user community, both within EPA and outside the Agency. The results of this analysis and the methodology employed in conducting it are presented in Appendix A; however, the salient information regarding the outside user community is summarized below. The percentages for each user group were derived from the sample analysis, with the number of searches extropolated to reflect an annual volume. This extrapolated annual figure of 620 requests differs considerably from the previous figures supplied by APTIC, presumably because of monthly variations in numbers of requests. In order to provide conservative estimates, the Project Team has used the high APTIC figure. In the table below, average turnaround is the average daily turnaround per search.

User Group	# Requests	# Searches	% of Total Searches	Avg. Turn-around (Days)
Other Federal Agencies	24	48	5.2%	21.8
State & Local Agencies	104	140	15.0%	25.0
Foreign Governments	24	20	2.1%	55.2
Industrial/Commercial	240	396	42.5%	26.1
Academia	92	156	16.7%	31.1
Private Citizens	24	40	4.3%	50.5
EPA	112	132	14.2%	21.5
TOTAL	620	932	100%	27.6

The above statistics indicate that turnaround is extremely slow, and this was substantiated in all of the user interviews conducted by the project. Every interviewee suggested improving turnaround as a means of improving the quality of service. Although APTIC functions to serve both EPA as well as the outside community, program management emphasized its desire that APTIC serve as a resource for those users who can not afford or do not have access to alternative sources of information. As the statistics above show, these user segments, Private Citizens and Foreign Governments, receive the least service in terms of turnaround. In addition their usage accounts for less than 10% of total non-EPA usage while Industrial/Commercial, State and Local Agencies, and Academia account for 86.5% of non-EPA usage. Our analysis of requests showed that 21% of the Industrial/Commercial sector were documented as performing literature searches as a direct result of government contracts or grants.

EPA usage of the system can be summarized by region in the following table:

EPA Regions	# Requests	# Searches	% of Total Searches	Avg. Turnaround
1	4	8	6.1	0.5
2	4	4	3.0	6.0
3	8	8	6.1	6.5
4	80	96	72.7	24.5
5	12	12	9.1	25.0
10	4	4	3.0	29.0
TOTAL	112	132	100%	21.5

Again the number of requests and searches have been extrapolated to reflect annual totals based upon our sample of user requests. The relatively quick turnaround for the first three users appears abnormal when compared to the last three. This apparent dichotomy can be explained due to APTIC's attempts to perform some searches on a rush priority and if the computer system functions properly, turnaround time can be minimized. The extremely large usage by Region 4 employees is not accidential since most of the requests sampled here were from the Research Triangle area where there is a high awareness of APTIC. Also much of the work performed by the Office of Air Quality Planning and Standards (OAQPS) is situated there. The variable usage by the other regions can be explained by many of the interviewees' comment that they had learned about APTIC very informally-usually through conversations with a co-worker. Regional employees who are separated from Research Triangle Park by large distances and who are only vaguely familiar with ATIC's services, are not likely to use it. The low usage by the EPA Regions and by those who can not afford other information sources is in conflict with management's stated objectives and indicates a need for more effort to increase both public and EPA awareness of the system.

The following table shows the level of detail for the searches picked in the request sample. These figures represent percentages of total searches rather than numbers of actual searches which should facilitate their interpretation.

	Level of Detail		
	General	Semi-General	Detail
EPA	3.4	5.6	5.2
Other Federal Agencies	-	4.7	0.4
State/Local	0.4	4.7	9.9
Foreign Government	0.4	-	1.7
Industrial/Commercial	5.2	23.2	14.2
Academia	3.4	4.3	9.0
Private Citizens	0.4	3.0	0.9
	13.2	45.5	41.3

In dividing these searches into the three classes of detail, the project used the following guidelines:

- General, including requests for information covering a broad subject without defining particular methodologies (i.e., "Air Pollution", "Atmospheric Interaction", etc.)

- Semi-Detail, which identifies an area within a larger subject category (i.e., "Atmospheric Interaction -- Aerosols")

- Detail, which specifically identifies a particular apsect of an area (i.e., "Atmosperic Interaction -- Aerosols; Light Scatter).

The table reveals that by far most users are submitting detailed and semi-detailed requests. Although the figures are not conslusive, they do indicate that perhaps APTIC's users are more cognizant of their needs (and therefore communicate more specific requests) than the users of some of the other EPA bibliographic systems.

(2) Operations Costs

The operations costs presented in this section reflect costs incurred during fiscal year 1975 and those budgeted for 1976. Because of the difficulty in obtaining the input Contractor's technical and cost proposals as well as other accounting information, some of our cost analysis is based upon extrapolated results from historical percentages. The costs which the project was able to get are presented on the following page.

Several comments should be made concerning how these figures were obtained. The aggregate costs for each of the five major categories, both estimated and actual expenditures, were obtained through the System Manager although the System Manager was not responsible for preparing the 1976 budget. The 1975 expenditure for Direct Operations ($114,000) was retrieved from the System Manager as actual expenditures; however, no detailed expenditure allocations were available. The 1976 Budget delivered to the Project Team merely divides Direct Operations into salaries and all other. The estimated figure for Printing ($69,000) was actually budgeted for Printing the previous year, but is believed by APTIC to also be a realistic figure for 1976. D.P. Associates, the contractor of the APTIC software system, is expected to provide approximately $25,000 in services by the end of the present contract, of which approximately $3,000 occurred prior to June 30, 1975, the end of APTIC's fiscal year. The allocation of the Franklin contract costs provided by manager who computed them from functional costs provided by the contractor in 1973.

Neither the cost proposals, the billings, nor the contract provide functional breakdowns of current costs. Data processing costs are budgeted for 1976 for $200,000; however, because the cost algorithm being used on the Univac

APTIC COSTS

SERVICE	FY 1975 EXPENDITURES		FY 1976 ESTIMATE	
I. Direct Operations:				
. Salaries			30,000	
. Printing			69,000	
. D.P. Associates			22,000	
. All Other		114,000	19,000	140,000
II. Franklin Input Contract:				
. Screening (12%)	45,261		45,360	
. Acquisitioning (14%)	52,805		52,920	
. Cataloging/Abstracting/Indexing (16%)	60,348		60,480	
. Editing (10%)	37,718		37,800	
. Keying/Proofing/Correcting (22%)	82,979		83,160	
. Miscellaneous/Cerical (4%)	15,087		15,120	
. Supervision (7%)	26,402		26,460	
. Materials & Fees (15%)	56,577	377,177	56,460	378,000
III. Japanese Literature Acquisitioning Contract		33,560		33,000
IV. APTIC Data Processing Costs		179,759		120,000
V. Other Costs Attributable to APTIC:				
. Document copies for Requestors		6,612		7,000
. Document copies from secondary sources		12,000		12,000
Total Costs		723,108		690,000

has recently been modified to reduce the costs of batch processing, the System Manager feels that $120,00 is a better estimate.

Finally, the Other Costs Attributable to APTIC are costs which are not absorbed by APTIC's budget but are expended from other budgets as a direct result of the APTIC service. The first item reflects the approximate costs (both actual and estimated) attributable to providing document copies to valid requestors (principally EPA employees and contractors). These costs were derived from APTIC's figures for the number of document requests they received last year, the average number of pages per document, and using a cost of $.05/page for reproduction. This cost is funded from the budget for the Office of Administration and is expected to be approximately the same for fiscal year 1976. The second item reflects the costs absorbed by the EPA Library in supplying Franklin with copies of documents which APTIC identifies as pertinent from scanning secondary sources. The above figures do not, however, reflect the increased contract costs which will result during fiscal year 1976 when APTIC begins delegating to the input contractor responsibility for processing requests and performing searches. For this reason and the conservative approach taken in estimating the above costs, the total of $690,000 for fiscal year 1976 is actually quite conservative.

Total costs alone, however, are not particularly useful in evaluating the system's performance unless they can be related to the services provided to users. Service is in itself both a subjective evaluation of the quality of the system outputs as well as a quantifiable measure of its usage. Without usage, a system can not possibly provide any service. In order to provide an objective measure of the cost of service provided, the following table compares the system's costs with usage in determining the cost per search. This is computed at the current level of usage and considers a doubling and tripling of usage.

	System Usage		
	Present	Double	Triple
. Fixed Costs (Total Cost minus the costs of copying documents and contractor software supporting of $10,000)	$673,000	$673,000	$673,000
. Variable Costs (reflecting increased volume)	-	17,520	26,280
. Total Costs	673,000	690,520	699,280
. Level of Requests for searching	876	1,752	2,628
. Cost/Request	$768.26	$394.13	$266.08
. Cost/Search (1.5 per request)	$516.50	$262.25	$177.03

The above figures are based on estimated total system costs ($690,000) for fiscal year 1976, which began June 30, 1975, and on the number of search requests (876) which APTIC has received from January through December, 1975. The fixed costs do not include the additional costs which will be incurred as a result of contract additions nor do they include $10,000 of the software contractor's fee which are estimated as the portion of contractor fees attributable to his efforts to debug the system. In addition, these figures also do not include the costs associated with additional job failures which would surely result from processing an additional workload on the Univac.

The variable costs, those costs which fluctuate with the volume of searches, were computed using an incremental cost of $10/request. This increment reflects the additional costs incurred as a result of preparing additional search strategies, keying and correcting additional input, processing a larger workload on the computer, ordering more supplies, increasing mailing costs, etc. Ten dollars per request is a conservative estimate particularly in view of the variable cost of $53 per search associated with SWIRS. Finally, the cost per search was calculated from statistics derived from our analysis of user requests. We found that for the period covered, each request generated on an average 1.5 computer searches. This figure was used in the cost per search calculation since the exact number of searches performed last year is not known.

These figures are meaningful in that they reflect the resources required to process various levels of demand as the system presently exists. Arthur Young & Company does not believe that these costs figures reflect acceptable levels of resource expenditures for the amount of services currently being provided. Even if the volume of requests were to triple, thus substantially reducing the costs allocated among users, the service would still expend $177/search (or $266/request). Arthur Young & Company believes that the service currently provided can and should be adjusted to provide a more cost effective product, as discussed in the RECOMMENDATIONS AND ACTION PLAN of this report.

(3) System Service

One evaluation of services provided by any system is that provided by the users themselves. Conversations with sampled interviewees, the analysis of search requests by various subject categories, and previous user surveys where they have been found to exist have provided valuable input to evaluating the services rendered by the EPA bibliographic systems.

It is important that each system attempt to periodically assess its service to determine if it serves the needs of the user community. To accomplish this function, a System Manager must be aware of the dynamic interests of the user groups and be able to relate this information to the selection and acquisition of new material for the data base. In this manner, a manager can dynamically alter his selection criteria to meet changing user requirements.

An overall assessment of APTIC's success in matching its acquisitioning to user requests can be made fairly easily since APTIC breaks its subject coverage into discreet categories and has developed statistics relating the coverage of each category to its total collection of documents. (Table I in the September 25, 1975 memo to Mr. Jean Schueneman, Director of the Control Programs Development Division, provides these statistics.) Because each document accessed by the input contractor is also provided in full hardcopy to APTIC, the document collection should closely parallel the contents of the data base. From the sample of search requests, a needs analysis can be performed in a straight-forward fashion which yields some interesting results when compared to the category breakdown of the data base.

Category	% of Data Base	% of Search Requests	Variance
Measurement Methods	15.0%	11.7%	+ 3.3
Emission Sources	9.2%	4.5%	+ 4.7
Effects-Plants & Livestock	4.0%	9.0%	- 5.0
Effects-Human Health	13.3%	8.1%	+ 5.2
Control Methods	26.1%	25.2%	+ 0.9
Basic Science Technology	9.0%	9.9%	- 0.9
Legal & Administrative	5.0%	0.9%	+ 4.1
Atmospheric Interaction	9.2%	4.5%	+ 4.7
Air Quality Measurement	4.5%	1.8%	+ 2.7
All Other*	4.7%	24.4%	-19.7

*This category includes all search requests which could not be readily classified as belonging to a specific category.

The large variance in the all other category is in itself not meaningful. It reflects the difficulty which the Project Team encountered in identifying a category interest from the search documentation and resulted in a large search statistic for that category. The maximum variance of the remaining categories is 5.2 which is relatively small. If we can assume that the large variance for all other can be evenly distributed among the remaining categories, these variances would become even smaller. These figures do indicate that APTIC's accessioning activities parallel user interests. A similar analysis of the inflow of document accessions and search requests performed on a periodic basis would indicate changes in user interest trends and measure whether the accessioning activity varies in response.

All of the users questioned in the user survey expressed high praise for the value of the service and its products. Of those questioned, however, only one-third were actual end users of the information. The remainder were librarians for businesses, OAQPS personnel obtaining information for contractors, regional employees satisfying requests for information from State and local agencies, etc. All information retrieved was to be used as 'background' information in particular subject areas, as the initial literature search for individuals in projects. Less than half of those questioned were aware of alternative sources of information such as those provided by commercial bibliographic services. Users consider the search results services to be of considerable value, but not absolutely critical to their projects. Appendix B contains additional data regarding the user survey undertaken by the project team.

Users considered the turnaround time to be excessively slow. The average turnaround for all user groups exceeded one month. This is particularly significant in view of a finding from a previous survey conducted by the Franklin Institute in the spring of 1973. That survey revealed that the majority of users were satisfied with five day turnaround but few regarded more than five days as acceptable. The Franklin survey interviewed 95 EPA employees in the Raleigh-Durham area and among other findings revealed that:

- The majority of EPA employees interviewed had little interest in Japanese literature or foreign or domestic patent literature (APTIC presently has a separate contract with the Franklin Institue for the acquisitioning of Japanese biomedical literature).
- The majority of EPA employees interviewed were interested in current year material only.
- 74% of the interviewees were not familiar with the APTIC Selective Dissemination of Information service.
- Most of the respondents thought that 25% of the search results were marginally and directly pertinent, and 50% of the results were nonpertinent.

- The majority of interviewees felt that having access to abstracts was very valuable.

Finally, two comments were voiced which are important and bear upon the quality of services rendered by APTIC. Three independent sources within the EPA Library reflected upon the poor quality of much of the document microfiche which APTIC has received from the Franklin Institute. Apparently the quality control over this function has improved in the past year, but many fiche are still of poor quality, resulting in poor quality reproduction.

In addition, it was suggested by two independent sources that the APTIC thesaurus is too broad and does not facilitate detailed searching. One commented, for example, that there is no synonym for "energy" in order to search for documents dealing with air pollution effects resulting from various means of generating energy. The thesaurus does appear to be rather nonspecific which tends to make searching of narrow "researcher oriented" questions difficult, particularly by those who are outside the system (i.e., the EPA librarians and information specialists). Other bibliographic systems that are commercially available and frequently queried by the EPA librarians, have more highly developed thesauri and are therefore relatively easy to search, even for a nonprofessional. Because of this difficulty, the librarians are unable to directly search the APTIC data base or even assist in the development of satisfactory search strategies. For these reasons and because of the slow turnaround, the EPA Library refrains from utilzing APTIC. If, indeed, APTIC wishes to increase its public awareness, it must develop a thesaurus which is useful and responsive to the needs of potential users. When this has occurred, a major obstacle to the expansion of further system usage will have been removed.

(4) Data Processing Characteristics

Prior to the development of the Technical and Environmental Information System (TENIS), APTIC utilized an IBM software package, Combined File Search System (CFSS), for performing all bibliographic searches. CFSS apparently performed well and adequately served as a bibliographic search system. Turnaround time on search requests using CFSS is reported to have been about five days. CFSS was abandoned only because the IBM 360 at Research Triangle Park was replaced with a Univac 1110, and the Data Systems Division (now a part of Management Information and Data Systems Division) felt that the costs required to convert the eighty-five assembly language programs comprising CFSS was greater than either the costs of developing a new system or purchasing a software package. A survey of appropriate systems was made by DSD in 1972, but the conclusion was reached that no close substitutes were operational which would not require extensive modification. Both System 2000 and DMS-1100 were examined, but it was felt that neither had existed long enough to demonstrate their reliability. As a result, DSD decided to design, develop, and implement their own system.

Although a number of existing search systems were examined prior to designing TENIS, the system was designed using the state-of-the-art techniques which had existed prior to 1972. Batch oriented systems were common, but the data base approach to developing applications was becoming popular. It was during this period that such on-line, interactive systems as STIMS/RECON were being developed and implemented but were not in wide use. These systems actually set the state-of-the-art as it related to sophisticated bibliographic processing.

Because it was necessary to develop TENIS before removal of the IBM hardware, a high priority was placed on quickly developing an adequate replacement for CFSS. For this reason TENIS was designed to fulfill the present needs of APTIC as they existed in 1972-1973 without a thorough examination of future requirements being made. The programming task was completed with about six man-months of effort and was done in a quasi-test mode. That is, the programmers were experienced in Cobol, the language in which TENIS is coded but had no experience using Univac hardware and software. As a result of this and the time constraint imposed upon the development group, the system was designed in a less than optimum fashion. The following processing inefficiencies have been identified by the development group and the staff of Arthur Young & Company:

- Inefficient Sorting Techniques

  The TENIS master files contain variable length records, and the system sorts variable length record files during both search processing and updating the data base. The Univac software expands each variable length record to its maximum size during sort operations consequently resulting in large amounts of direct access work areas being used during sort processing.

- Excessive Search Requirements

  When processing a batch of searches, the system must pass the entire master file which consists of eight reels of magnetic tape and requires approximately two hours of wall clock time. During this processing, each record on the master file is examined to determine if its accession number matches a search strategy. Our user survey as well as previous surveys have indicated that the majority of users are interested in current year material only. For these searches as well as the Selective Dissemination of Information service provided by APTIC, only a relatively small portion of the master file needs to be processed. Some consideration should be given to segmenting the master file so that a relatively small subset of it can be accessed to satisfy those requests for current literature only.

- Redundant Searching Performed for On-Line Requests

  The APTIC search system has an on-line capability allowing requestors in the various regions to query the data base from terminals in an interactive mode.

  However, only bibliographic citation information and not the abstracts, are available on-line. Frequently when these users have identified a set of citations which interest them, they request abstracts of the corresponding documents. In order to obtain these abstracts, they must telephone APTIC and request that a batch search be made which matches their own strategy or the accession number of those abstracts which they have chosen on the basis of the citations which they reviewed interactively. The current system does not allow regional users to automatically trigger the printing of abstracts from their terminals.

- Restructuring the Job Stream

  The project team has observed that several batches of jobs are run through the system daily. APTIC personnel and the contractor stated that they divide the searches into batches rather than running them all in one job to avoid space overflow occurring on the various files. Although running searches separately does require less file space, it also requires multiple passes of the master file. Since this file consists of eight reels of magnetic tape and requires approximately two hours to read, multiple passes dramatically increase processing time and costs. Even jobs performing a few searches must presently read the entire master file and thus use an inappropriate amount of resources. However, since processing the master file is I/O bound, doubling the number of searches in a job would not significantly increase processing time. The system manager also stated that dividing the searches into multiple jobs was done to increase the probability that some searches would get processed. APTIC has experienced so many job failures during the past year that it has developed a substantial backlog of searches. However, this is an expensive means of avoiding problems which could be solved by concentrating additional programming effort on debugging the system. This point is addressed in greater detail below.

- System Malfunctions

These problems are attributable to program bugs, operator errors, and hardware malfunctions. All three types of problems have plagued APTIC since the computer system was implemented. From November 1 through December 24, 1975, fifty-seven jobs have failed as a result of these problems, and consequently APTIC has developed a large backlog of searches. Appendix D provides a complete list of these problems which have been maintained by APTIC since January 1.

Since APTIC personnel are not data processing oriented, it is difficult to determine whether many of the problems are operationally-oriented or software oriented. However, an indication of the type and frequency of errors can be obtained. Frequently cited problems include:

- Maximum job time exceeded
- Storage overflow
- Parity Errors and hardware checks
- File entry dropped from Master Directory
- Operator cancelled job backlog

In addition, several comments were made by the APTIC people and their contractor which bear upon these issues and further indicate the nature of these problems. These are as follows:

- In performing on-line searches of the citation file, the number of documents satisfying a strategy is displayed on the user terminal. Frequently this number is not correct.

- Several combinations of subject terms have been identified which, when entered into the system, cause the program to enter an apparent "endless loop". Eventually the maximum run time parameter is exceeded in these situations and the job is cancelled.

- Program documentation is not as complete as it should be. In particular it should include program listings, flow charts, complete descriptions of the input/output files, descriptions and resolutions of error conditions and codes, and a comprehensive description of program processing. APTIC personnel have stated that frequently error conditions occur for which neither the cause nor the solution is readily identifiable. On several occasions, there has been confusion as to which program listing matches the current operational program version. Formal standards and procedures should be followed to ensure that control is exercised over the modification of programs and their documentation. This is particularly important where contractors are involved in maintaining software systems as is the case with APTIC.

The development group has estimated that approximately 6 to 12 man-months would be required to optimize the APTIC search system. This represents 100%-200% of the initial programming effort; however, whether or not the present system is optimized to its fullest, consideration should be given to at least correcting those program bugs and operational problems that significantly contribute to downtime.

(5) Source Overlap

The Analysis of Required Journals, Appendix C, describes the analysis performed on the required journals of each of the Technical Literature Systems. However, a substantial overlap exists between those journals whose screening is required by APTIC, those which are required by SWIRS and Pesticides, and those which are normally screened by the various commercial abstracting services. The following table illustrates the percentage of the APTIC required journals also used as sources by the other services reviewed:

SERVICE	PERCENTAGE OF APTIC JOURNALS REVIEWED
CHEMICAL ABSTRACTS (CA)	81.9%
ENGINEERING INDEX (EI)	18.5%
INDEX MEDICUS (IM)	28.9%
CA OR EI OR IM	90.0%
SWIRS	35.3%
NOISE	21.1%
PESTICIDES	37.1%

The fact that a required APTIC journal is reviewed by another service does not prove that the same individual articles are selected. However, since there is such a great degree of overlap, there is a definite possibility of a considerable degree of overlap in selected documents, particularly since Chemical Abstracts, Engineering Index, and Index Medicus do include APTIC-related subjects in their selection criteria. In addition, the high degree of source overlap among the required journals of the EPA systems suggests that a combined screening and acquisition activity would be beneficial.

The Project Team selected 11 actual searches at random to also be run on commercial systems, as a comparison to data contained on the APTIC data base with that available commercially. Time and cost considerations forced the Cincinnati Library, the organization performing the searches, to perform only three of the selected searches. The results of these three searches included very little duplicate articles. A sizable number of hits were received on two queries of the APTIC system but relatively few on the commercial systems. However, from Project Team's observation the APTIC searches were so loosely structured as to select literally anything remotely connected to the original subject request. Selection criteria did not single out the subject. For example, a search on air pollution effects on cotton produced abstracts relating to growing cotton as well as cotton fiber filters and dying cotton fibers. At least one duplicate abstract was noted within the original APTIC search but assigned a different citation number, and a significant number of author's names were not in proper format. Normally, last names should be first, followed by first name and middle initial. A great many instances were noted where author was with given name (or initial) first followed by middle initial and last name. As discussed in the RECOMMENDATIONS AND ACTION PLAN Section, these results dictate that caution be used in any assumption of the existence of parallel or duplicate contents of data bases.

	NUMBER OF CITATIONS SELECTED					
SEARCH SUBJECT	APTIC	MEDLINE	BACK 72	BACK 69	BACK 66	DUPLICATES
Air Pollution on Lipid Metabolism	177	12	1	9	1	8

	APTIC	CA	EI	NTIS	DUPLICATES
Acrylomide, Area Formaldehyde and Urthean and Soil	4	57	1*	4*	0

* Copies of abstracts not received.

	APTIC	BA	CA	CAIN	NTIS	POLL. ABSTR.	DUPLICATES
Air Pollution Effects on Cotton	196	4	3	2	1	3	5

Note: There were five duplicated citations. Of the five different APTIC articles present on commercial systems, two appeared on more than one commercial system.

4. COMMERCIAL ABSTRACTING SERVICES

The use of commercially abstracted and prepared data bases, accessed through a commercial data processing service would substantially reduce the cost borne by APTIC. This would be true even if APTIC were to continue its selection and abstracting activity at a level equal to half its current effort (to include documents not now selected commercially) and donate these selections to a commercial data base. In fact, it is likely that a commercial service would agree to expand its selection coverage to include these items in return for the current APTIC data base. The following analysis portrays the costs which would be borne by APTIC under the following three conditions:

- No usage of commercial systems
- Total reliance on commercial systems, but with continued selection and acquisition activity equal to half of the current effort. The resulting data base would be donated (or sold) to one of the commercial abstracting services.
- Total reliance on commercial systems.

Note: Chemical Abstracts, for example, has expressed interest, as a condition of receiving the current EPA data bases, in expanding its own selection activity to include the EPA criteria.

		COMMERCIAL SEARCHING	
COST COMPONENT	NO COMMERCIAL USAGE	HALF CURRENT ACCESSIONS	NO ACCESSIONS
APTIC COST			
Salaries	$30,000	$30,000	$30,000
Contractor Acquisition and Input Cost	377,177	188,588	- - -
Data Processing Cost	179,759	59,919	- - -
TOTAL APTIC COST	$586,936	$278,507	$30,000
COMMERCIAL SEARCH ($50/search)	- - -	40,200	40,200
TOTAL COST	$586,936	$288,707	$70,200

Even if APTIC usage were to triple to 2412, savings would still be substantial:

		COMMERCIAL SEARCHING	
COST COMPONENT	NO COMMERCIAL USAGE	HALF CURRENT ACCESSIONS	NO ACCESSIONS
APTIC COST			
Salaries	$30,000	$30,000	$30,000
Contractor Acquisition and Input	377,177	188,588	- - -
Data Processing Cost	187,799	125,199	- - -
TOTAL APTIC COST	594,976	343,787	30,000
COMMERCIAL SEARCH ($50/search)	- - -	96,480	96,480
TOTAL COST	$594,976	$440,267	$126,480

The following factors have impacted the figures used in the above analysis:

- The only figures depicted in this analysis are those related to operations critical to the performance of searches. Costs not involved were printing, D.P. Associates Contract, other directed operations, Japanese Selection and Copying, for a total of $136,172.

- Data processing costs for a tripling of usage was computed by increasing 1975 data processing cost by $10 for each additional search.

- The cost allocated to a search of commercial data bases is taken from independent research performed by Mrs. Mary Calkins, Information Retrieval Librarian at the EPA Cincinnati Library. The methodology and results of this research is currently in preparation for publication. This research, performed using actual searches performed by the Cincinnati Library, resulted in an estimated cost of $15.68 per search on any single commercial system (free government-operated systems were not included in the study). During the period examined, an average of 2.7 data bases were queried through any single data processing system, using an average of 20 minutes connect time. Additional costs, including terminal costs and telephone costs add $4.63 for a total search cost of $20.31. The Project Team has doubled this figure (to $40/search) in order to ensure a conservative comparison.

5. PRIMARY CONSIDERATIONS

It would seem that there would be a strong National interest in an air pollution information retrieval system, particularly in view of the recent ecological interest at the National, State and local levels. However, actual usage of APTIC peaked in 1972 and has decreased by 50% since that time. Considering a current cost of $723,108 and annual usage of 876 requests, Arthur Young & Company believes APTIC's cost justification should be reviewed. While part of this decline may be due to the serious data processing problems APTIC is experiencing, it is difficult to ascribe all of the decline to poor turnaround, particularly since the decline began before the implementation of the new system. Finally, while APTIC strives to serve those who cannot acquire such information elsewhere, only a small minority of users fall into this category. Private citizens, Citizen Action Groups and foreign governments together received only 60 searches or 6.4% of the total.

If the Air Pollution Technical Information Center is to fulfill the mission which it has assumed for itself, it must vastly increase the number of users benefiting from the system, particularly those without an ability to acquire this data elsewhere. If the amount of effective usage of the system cannot be increased there is little justification for its continued existence. Since the funds and effort needed to increase usage of the system must come from elsewhere in the APTIC service, it seems unavoidable that acquisition and/or data processing expenditures must be reduced. The ANALYSIS OF REQUIRED JOURNALS, Appendix C, performed by the Project Team, indicated that a considerable portion of the APTIC data base may be available elsewhere. In addition, commercial services have expressed interest in maintaining this type of data base if presented with the current data base. Considering the above factors, the tremendous decrease in costs which would result from a greater reliance upon commercial services and the problems being encountered with the APTIC data processing system, the Project Team believes a reallocation of emphasis from selecting, abstracting and acquisitioning documents to publicizing the system and providing more responsive service is mandatory. The steps to be taken in accomplishing these goals are discussed in the RECOMMENDATIONS AND ACTION PLAN.

V. PESTICIDES TECHNICAL/LIBRARY RESEARCH SYSTEM

The Pesticides Technical/Library Research System provides for the dissemination of literature concerning the effects of pesticides upon non-target organisms. The Pesticides System does not serve the public, but exists to service the Office of Pesticide Programs (OPP) of the EPA. Although the service will generally provide searches on request to other EPA and Federal Government users, no effort is made to publicize the system beyond the OPP. Requests from other than Federal Government sources are normally transferred to the Office of Public Affairs or returned with an identification of persons or organizations active in the subject area identified by the request, although preprinted bibliographics on the subject will be sent to the requestor if they already exist. The Office of Pesticide Programs believes that there would be a public interest in the system should it be publicized, but does not have sufficient funds to support this potential user population, nor is such activity considered to be the responsibility of OPP. The Pesticides data base is on the TOXLINE System of the National Library of Medicine, and can be utilized by any user who has access to that system.

1. SYSTEM DESCRIPTION

This discussion presents a brief review of the manner in which the Pesticides System responds to the operational requirements of the Office of Pesticide Programs. A more detailed description can be found in the EPA TECHNICAL LITERATURE SYSTEM PROFILES, produced separately under this Task Order.

Under contract to the Environment Protection Agency, the Franklin Institute selects documents and articles from the world's literature for inclusion in the Pesticides System, using selection criteria developed by the Technical Services Division. The contractor enters citations for selected documents onto computer tape through a DATAPOINT 100 minicomputer using an interactive program providing a limited amount of on-line editing. This tape is used to generate a listing of citations to be sent to the Technical Services Division, and is also marketed by the contractor to the National Library of Medicine for use on its TOXLINE System. As the contractor abstracts the selected documents, xerox copies of the document are sent to the Technical Services Division. The abstracts themselves are used to create a camera negative of the Pesticides Abstract Bulletin, which is provided to the Government Printing Office for generation of the bulletin.

The activities performed by the Technical Services Division of the Office of Pesticide Programs are described in the EPA TECHNICAL LITERATURE SYSTEMS PROFILES. Primary data creation activities include:

- Use of the 'Cite List' to generate a current awareness list to be provided throughout OPP.
- Generation of microfiche copies from xerox copies of selected articles
- Review of contractor quality control through parallel review of 60 'Core' journals.

The Technical Services Division is in the process of developing its first automated bibliography system, patterned in concept after that being used by the EPA Library. The system is a batch oriented sequential file system which, while capable of performing automated searches, will primarily be used to generate Computer-Output-Microfiche copies of the accumulated Citation List in various keyword sequences. Rather than using a keyword thesaurus, the system will apply the numeric hierarchic classification codes utilized for the manual Pesticides System previous to 1969. All citations selected since 1969 have been reviewed by the contractor, and classification codes have been assigned in addition to the thesaurus keywords. The system shall consist of two primary data bases, the Retrospective (pre-1969) and Current Awareness (post-1969) Files. These files are to be identical in format but maintained separately since most search requestors are interested only in current literature.

In performing searches for the EPA, the Technical Services Division examines not only its own, currently manual, data bases but also those of the major commercial abstractors active in the area of Pesticides. These include such commercial and government data processing systems as those of Lockheed and the National Library of Medicine.

2. MANAGEMENT PLANS AND OBJECTIVES

The Management Plans, Needs and Objectives of the Office of Pesticide Programs which are supported by the Pesticides Technical/Library Research System have been documented in the 'PROGRAM MANAGEMENT OBJECTIVES REPORT' produced under this same Task Order. The primary factors noted by management are discussed below:

(1) Program Management

Program Management has been pleased with the performance of the Pesticides System, and looks for this system to provide considerable assistance in the OPP effort to register and classify over 40,000 products by October 1976. However, there exist no detailed objectives, performance requirements, cost/benefit analyses or user need evaluations which would enable management to effectively evaluate system performance in detail. In addition, while a study was made of the resources available to OPP in fulfilling the registration effort, no determination of the detailed technical information requirements of this registration activity has been communicated to the Technical Services Division.

(2) System Reliability

Because of the need for fast reliable fulfillment of user requests, particularly when generated during hearings or litigation, management is reluctant to rely on external systems as its sole means of information gathering. This concern is discussed in the RECOMMENDATIONS AND ACTION PLAN of this report.

(3) Scope of Coverage

The Office of Pesticide Programs is expanding its scope of interest from simply the health aspects of pesticides to such areas as Integrated Pest Management and the economic impact of developments in the pesticides area. In addition, program activity is moving into such areas as Methods Development for Quality Assurance and the development of a National Monitoring Program. These changes may vastly expand the scope of literature to be included in the Pesticides System. While these potential activities do not have a direct impact on the EPA Technical Literature Systems Analysis, they do make more critical the need to provide technical literature services in the most economic method practical.

(4) Data Systems Development

Management is interested in both the use of the Pesticides data base to support an interlibrary loan system and the use of the Pesticides software as a starting point in developing a Company Data Storage and Retrieval System, providing information regarding pesticide manufacturers and their products or processes. While any recommendations involving change to the Pesticides operating environment must consider these requirements, a substantial savings in the system itself would outweigh the benefits of such parallel support.

(5) Minicomputers

Management is interested in the use of minicomputers as a means of reducing operational complexity and cost, and of increasing flexibility and control. While the subject of minicomputers cannot be addressed in depth during this study, some considerations are discussed in the RECOMMENDATIONS AND ACTION PLAN, Appendix G.

3. ANALYSIS OF OBSERVED CHARACTERISTICS

This discussion addresses those characteristics which the project team considers to be most germane to the purposes of this study or in which particular improvements can be made. Where appropriate, comments are made concerning the impact of particular features upon overall system performance, and recommendations set forth which will improve performance or reduce costs experienced. The characteristics noted in this section are also used in supporting discussions in the RECOMMENDATIONS AND ACTION PLAN.

(1) System Usage

During 1975, the Technical Services Division (TSD) responded to approximately 340 search requests, of which only 38 originated from non-EPA users. In addition, the OPP Criteria and Evaluation Division (C&E) performed approximately 250 searches completely independently, using data bases supported by Lockheed. The searches performed by the Technical Services Division were calculated from the January-June portion of the TSD Request Log kept by that Division and extrapolated to annual values. The figure for independent Criteria and Evaluation activity derived from an estimate made by the C&E Information Office.

The composition of the requestors served by TSD includes:

REQUESTOR	NUMBER OF REQUESTS	
OPP(DAA and Policy and Operations Staff)		4
OPP Criteria and Evaluation Division		236
Chemical	58	
Plant Studies	54	
Metabolic Effects	42	
Economic Effects	10	
Ecological Effects	16	
Administration Hearing Staff	6	
Information Office	50	

OPP Registration Division		26
Chemical	18	
Herbicide/Fungicide	2	
Insecticide/Rodenticide	4	
Toxicology	2	
OPP Operations Division		4
OPP Technical Services Division		6
OPP Strategic Studies Unit		8
Total OPP		284
Other EPA		18
Office of Water Programs	2	
Regional Offices	10	
Office of Research and Development	6	
Other Federal Government		14
USDI	4	
USDA	6	
Navy	2	
USGS	2	
Other		24
State Agencies	4	
World Bank	4	
Private Citizens	12	
Other	4	
Total Non-OPP		56

Since TSD responds to queries with information gathered both from the Pesticides data base and those commercial data bases available through Lockheed, users are normally not concerned or knowledgeable about the source of the information provided. In fact most users interviewed were not even aware that the Pesticides System exists; they simply know that if they request information of the Technical Services Division (or the Criteria and Evaluation Information Office) it will be provided. Users generally make extremely broad requests, partially because of the concern that more narrowly defined requests will result in the exclusion of something pertinent. Of the 166 requests to TSD reviewed by the Project Team, only 35 could be classified as detailed or even semi-detailed requests. Interestingly, the C & E Information Office has stated that they normally process detailed requests themselves, without utilizing the Pesticides System at all.

The Technical Services Division anticipates a substantial increase in number of searches. The cause of this future requirement is that the Registration Division is registering 40,000 products. TSD expects these requests to be quite detailed to support the "Rebutable Presumption" technique to be used in evaluating these products. Although a "Strategy for Controlling the Adverse Effects of Pesticides" has been developed for this effort by the Office of Pesticides Programs, the project team has been able to locate no detailed estimates of the type and quantity of information which will be required of the Pesticides System.

(2) Operations Costs

The cost and budgetary information presented in this discussion represents the calendar year 1976, since these were the most reliable figures the project team was able to obtain. Total costs attributable to the Pesticides System include:

SERVICE	EXPENDITURE
Normal Operations	
Technical Services Division Salaries	$122,193
Franklin Institute Selection, Acquisition, etc.	165,084
Data Processing Costs	16,504
Pesticides Abstract Bulletin (GPO)	13,000

Medline, Toxline	2,916
Lockheed	3,857
Total Operating Costs	323,554
System Development Activities	
Sigma Data System Development	35,000
Franklin Data Conversion Contract	20,000
Data Processing Costs	15,504
Total Development Cost	70,504
Total Pesticides System Cost	$391,142

Technical Services Division salaries were computed by multiplying salaries for each position by the proportion of time devoted to the Pesticides System. The Franklin Institute contract expenditure was calculated by prorating the expenditures for January through October to create an annual figure. Data processing costs for Fiscal Year 1975 (at NIH) were allocated in equal portion to normal operations and development of the new data processing system. The cost of the Pesticides Abstract Bulletin was taken from the OPP Fiscal 1976 estimates. Costs for the Sigma Data and Franklin Institute data conversion contracts were estimated by the System Manager.

The above figures were used to determine the overall cost/search of the system. In performing this analysis, total cost was computed by subtracting from 'Total Operating Costs' above, the costs of the Pesticides Bulletin, NLM and commercial searches, and the salaries of the librarian and librarian's assistant, since these services are not critical to the performance of searches. This results in a total search cost of $285,283. The computation provides cost/search for three eventualities:

- Under current usage
- Assuming tripling of usage (due to Registration activities)
- Tripling of usage and performance of all C&E searching

NUMBER OF SEARCHES	COST/SEARCH
Current (340)	839.07
3 x Current (1020)	279.69
3 x Current and C&E (1270)	224.43

(3) System Service

The Technical Services Division appears to be generally responsive to expressed user needs. User interviews, described in Appendix B, indicate overall satisfaction with search products and turnaround time (an average of 10.6 days). However, both the Project Team and System Management believe the requests received by the system to be too general in scope to receive the quality of service which such a system should provide.

Currently searches are made through computer query of TOXLINE, MEDLINE and those data bases serviced by Lockheed, and users receive abstracts generated through these systems. Once the new Pesticides data processing system is operational, users will be provided abstract numbers and will be expected to locate the abstracts themselves in the Pesticides Abstract Bulletin.

(4) Data Processing Characteristics

Since the planned Pesticides data processing system is not yet implemented, the project team was unable to make an evaluation of operational efficiency. The concepts and planned operations of the system have been reviewed and appear viable. However, the system's success depends on one central characteristic; the usage of the Classification Codes.

As discussed in the description of SWIRS, batch data processing bibliographic systems are considerably inferior to online systems because of their inability to provide the user with an immediate estimate, and examples, of the number and type of citations which a particular query will retrieve. In addition, sequential bibliographic systems are much less efficient than direct access systems, since an entire file must be accessed rather than only those records possessing the desired characteristics.

In the case of the Pesticides System, management believes that they have avoided these problems by the simple fact that searching will be performed not by a computer, but manually, using microfiche listings generated in various sequences from the master file through Computer-Output-Microfiche (COM). These microfiche files are nothing more than a microfiche library catalog, and their utility rests completely on the usefullness of the keywords on which they are sequenced. If a user is interested only in documents which address all of two or more subject terms and if these terms are not related in a specific hierarchy and in that sequence on the catalog listing, the listing cannot provide the necessary definition. It was to overcome this limitation that fully computerized bibliographic systems were developed, and are utilized elsewhere.

If the hierarchy of the Pesticides Classification Codes is sufficiently well-designed (and the classifying process sufficiently well-excuted) to provide this definition and correlation, the microfiche system will be effective. If not, serious problems in both efficiency and service may result. While the project team is unable to make an evaluation of the system, several knowledgeable and disinterested parties have termed the Classification scheme excellent in the hands of knowledgeable users. Since the microfiche copies are to be provided to the regional libraries, it is assumed that considerable training will be necessary before regions will be able to effectively utilize the systems.

Finally, it was determined that no cost or feasibility study was undertaken before the development of this system. Since the EPA Library appeared satisfied with their own version of this type of system, OPP assumed it would be effective for them also. Had the system been developed using inverted files and a generalized data management system, OPP would have had the option to go to efficient computerized searches if the microfiche system was not as viable as anticipated.

The Project Team did note that the work flow necessary to input and maintain the Pesticides System is cumbersome and inefficient. This does not cause particular concern since it is a carryover of operational methods used at a previous facility in Chamblee, Georgia, is recognized as such by management and is being modified. A brief description of the primary problems is included for completeness:

- Although the selection and acquisition contractor utilizes an automated file to generate the list of selected citations, TSD must re-enter that list through WYLBUR in order to create its Current Awareness List. Plans are being made to accept the tape directly from the contractor to be converted by computer program to the format required by TSD.

- Sixty journals are currently received both by TSD and the contractor, as a means of quickly gathering current awareness citations and of reviewing the selection of citations by the contractor. The value of this activity is currently being reviewed by TSD. The Project Team recommends its discontinuence, since:
  - It is doubtful that the increased timeliness is that great and, if it is, steps should be taken to accelerate contractor performance.
  - Since the contractor is aware of the 60 journals which are screened, its value as a quality control effort is slight.
- The effort required for microfilming copies of selected articles is both tedious and cumbersome in its present design, as described in the EPA TECHNICAL INFORMATION SYSTEM PROFILES. This effort will be included in the next TSD Request for Proposals or, if not included in the eventual contract, will be redesigned by TSD.

(5) Source Overlap

The Analysis of Required Journals, Appendix C, describes the analysis performed using the required journals of each of the Technical Literature Systems. A substantial overlap exists between those journals whose screening is required by Pesticides, those required by SWIRS and APTIC, and those which are normally screened by the various commercial abstracting services. The following table illustrates the percentage of the Pesticides required journals also used as sources by various other services:

SERVICE	PERCENTAGE OF PESTICIDES JOURNALS REVIEWED
Chemical Abstracts (CA)	70.5%
Biological Abstracts (BA)	58.0%
Index Medicus (IM)	32.7%
CA or BA or IM	81.9%
SWIRS	25.0%
APTIC	9.9%

The fact that a required Pesticides journal is reviewed by another service does not in itself prove that the same individual articles are selected. However, since there is such a great degree of overlap, there is a definite possibility of substantial overlap in selected documents, particularly since these commercial services (and others such as MEDLINE) do include some Pesticides-related subjects in their selection criteria. In addition, the high degree of overlap among the required journals of the EPA System suggests that a combined screening and acquisition activity would be beneficial.

The project team had planned to examine 10 random actual searches performed by the Technical Services Division to estimate the amount of overlap in citations contained in Pesticides and various commerical systems. Because of the voluminous responses to the very general requests submitted to the Technical Services Division, this effort was adjudged to be too time-consuming. One of the smaller returns was partially reviewed, a request for information concerning Dichlorophenoxyacetate, Dichlorophenoxyacetic Acid or 2,4-D. The amount of information delivered to the user included.

INFORMATION	NUMBER OF ABSTRACTS
Biological Abstracts	884
Chemical Abstracts	429
Excerpta Medica	59
MEDLINE (Excluding Pesticides)	370
MEDLINE (Pesticides Only)	62
NTIS	22

Of 27 of those articles contained in the Pesticides data base, 15 were duplicated by at least one of the other services.

6. COMMERCIAL ABSTRACTING SERVICE

Substantial cost savings could be realized through the use of commercially developed data bases supported by external data processing systems. The following anaylsis portrays the costs which would be experienced through such an approach. In performing this analysis we have assumed only those costs critical to the search effort, eliminating salaries of library staff and the system manager, the Pesticides Abstract Bulletin, one half of data processing costs and all current costs of external data bases. The cost columns depict anticipated costs for:

- No usage of commercial systems
- Total reliance or commercial systems, but with continued selection and acquisition activity equal to half of the current effort. The resulting data base would be donated (or sold) to one of the commercial abstracting services.
- Total reliance on commercial systems:

NOTE: Chemical Abstracts, for example, has expressed interest in acquiring EPA data bases and expanding their own selection activity thereafter to continue their support.

COST COMPONENT	NO COMMERCIAL USAGE	COMMERCIAL SEARCHING HALF CURRENT ACCESSIONS	 NO ACCESSIONS
TSD SALARIES (91,478)	$ 91,478	$ 80,328	$ 71,478
CONTRACTOR ACCESSIONS	165,084	82,542	-
DATA PROCESSING COSTS	12,000	6,000	-
TOTAL PESTICIDES	$268,562	$168,870	$ 71,478
COMMERCIAL SEARCH ($40/SEARCH)	-	13,600	13,600
TOTAL COST	$268,562	$182,470	$ 85,078

Even if system usage should increase by a factor of 4, to 1,360, the cost savings would still be substantial:

TSD SALARIES*	$ 91,478	$ 80,328	$ 71,478
CONTRACTOR ACCESSIONS**	165,084	82,542	-
DATA PROCESSING COSTS	12,000	6,000	-
TOTAL PESTICIDES	$268,562	$168,870	$ 71,478
CONTRACTOR SEARCH ($40/SEARCH)	-	54,400	54,500
TOTAL COST	$268,562	$223,270	$125,878

* Any increases in salary should be identical for each alternative.

**It is likely that such a large increase in usage would involve a concurrent increase in scope of literature covered.

The above estimates required the following assumptions:

- Since no cost figures are available for the new data processing systems, this figure can only be roughly estimated. A total data processing cost of $1,000 per month has been estimated, assuming the following monthly operations on prime computer time.
  - Two batch edit/updates
  - Five batch searches
  - Five microfiche generation runs (for five different sequences)
  - One current awareness program run
  - One Mailing Label generation.
- This analysis has assumed that if one half of the data acquisition activity is eliminated, TSD salary can be reduced by $11,500; and if all acquisition activity is eliminated, TSD salaries can be reduced by an additional $8,500.
- The cost allocated to a search of commercial data bases is taken from independent research performed by Mrs. Mary Calkins, Information Retrieval Librarian at the EPA Cincinnati Library. The methodology and results of this research is currently in preparation for publication. This research, performed using actual searches performed by the Cincinnati Library, resulted in an estimated cost of $15.68 per search on any single commercial system (free government-operated systems were not included in the study). During the period examined, an average of 2.7 data bases were queried through a single data processing system, using an average 20 minutes connect time. Additional costs, including terminal costs and telephone costs add $4.63 for a total search cost of $20.31. The Project Team has doubled this figure (to $40/search) in order to ensure a conservative comparison.

5. PRIMARY CONSIDERATIONS

The avowed purpose of the Pesticides Technical/Library Research System is the provision of needed services to the various divisions of the Office of Pesticide Programs. However, the Project Team must question whether this product is worth $839.00 per search to the EPA, particularly when commercial systems provide similar information for substantially lower cost.

The OPP has never performed a detailed examination of its information needs nor, apparently, is one contemplated for the upcoming registration activity. The type of information requests processed to date is also of concern to the Project Team. Of 166 requests examined during the analysis of sample requests (Appendix A), only 35 could be considered even moderately detailed. The others should, in fact, have been satisfied by the requestor himself through a manual review of published abstract bulletins. This not only indicates a surprising lack of sophistication on the Part of users, but also brings into question OPP's assumptions concerning the specificity and inclusiveness of user needs.

An additional area of concern is the duplication of function being performed by TSD and the OPP Criteria and Evaluation Division. The Project Team can see no justification for the maintenance of a separate search and document retrieval capability within C & E. While the physical separation of the two divisions may be a hindrance, methods should be developed to ensure specificity on requests submitted. The practice of funneling all C & E requests through the C & E Information Office virtually ensures a lack of communication between the requestor and a TSD search specialist. Finally, if delays in receiving information are serious, efforts should be taken to improve TSD service, rather than the creation of a parallel search capability.

THE RECOMMENDATIONS AND ACTION PLAN of this report contains a further discussion of these problems with recommendations for their resolution.

VI. EPA LIBRARY SYSTEMS

Each of the Regional EPA libraries is organizationally located within the operating structure of the Region which it supports; the Library Systems Branch acts as a coordinating force among all of these libraries. The Library Systems Branch develops systems and provides services which any individual library may utilize or not utilize as it sees fit. In spite of the potential difficulties inherent in such an organizational framework, the Library Systems Branch and the libraries themselves have succeeded in effectively cooperating, at least in those areas covered by this study.

1. SYSTEM DESCRIPTION

Of the four systems included in the EPA Library Systems examined during this project, only three are actually within the purview of the Library Systems Branch:

- The EPA Reports System, providing citations and abstracts of all formal reports generated by the EPA or its contractors
- The International Environmental Reports System, containing citations and abstracts of all documents provided by foreign governments through the International Documents Exchange Program and considered germane by the Office of International Activities.
- The Document Control System, containing citations only of all documents acquired by any of the individual libraries.

The fourth system, the Air Ecology System, while sharing the same general systems design as the other systems, is actually the property of the EPA National Ecological Research Laboratory in Corvallis, Oregon. While the Air Ecology computer requirements are currently funded by the Library Systems Branch, it is anticipated that this responsibility will be transferred to the National Ecological Research Laboratory. For the purposes of this discussion, we shall use the term 'EPA Library Systems' to address the three systems actually maintained by the Library Systems Branch.

The EPA Library Systems are unique among the EPA Technical Literature Systems, including the Air Ecology System, in that no formal selection and acquisition activity is performed and no attempt is made to provide a comprehensive data base of all literature applicable to a particular subject area. In fact, even though abstracts are included for the first two of the Library Systems, this function appears to be more that of an automated library catalog system than a technical literature search system and, as such, they exist in concert with the other library catalog systems, the Journal Holdings Inventory and the Hard Bound Book System. The Library Systems appear to have been developed primarily because of the need to catalog library holdings; the other EPA Literature Search Services have developed libraries primarily as an additional service to the users of their search systems.

The four services each utilize simple sequential batch data processing systems, quite similar in design to one another. A master file of citations is maintained for each of the systems and periodically sorted into various keyword sequences to produce the microfiche listings. They are also used to produce specialized subject or geographic bibliographies on request. New citations are added to the file using WYLBUR and an IRS merge program. In addition to the Citation File, the EPA Reports and International Environmental Reports Systems also contain abstracts on tape files, which are used to generate microfiche abstracts.

Detailed descriptions of each of the EPA Technical Literature Systems are supplied in the EPA Technical Literature Search SYSTEMS PROFILES, developed separately under this task order. The primary characteristics of each system are discussed below:

(1) EPA Reports System

The EPA reports system is operated by the Headquarters, EPA library to provide bibliographic information on all reports produced by EPA programmatic areas, their contractors and their grantees. Abstracting, citation and indexing efforts are performed by the programmatic areas themselves, while the generation of the primary input file is performed by the National Technical Information Service (NTIS), through an interagency agreement. The EPA library system utilizes this input file to create its own EPA Reports File and microfiche citations. In addition, NTIS provides microfiche copies of all EPA hard-copy reports, and a quarterly bibliography to all EPA libraries. Microfiche copies of hard-copy EPA reports are also sent by NTIS to countries possessing information exchange agreements with the EPA. The EPA Reports File currently contains approximately 8,000 citations, with a growth rate of 2,000 per year.

(2) International Environmental Reports System

This system was developed for the Office of International Activities to facilitate the administration of the International Documents Exchange Program. Participating countries exchange government reports relating to environmental issues. The Office of International Activities receives foreign documents and disseminates domestic reports to other countries. Those reports which are judged by them as having lasting value are sent to the Library of Congress where translations of non-English documents are made. The Library of Congress prepares citations and indexes, which are input to the International Environmental Report System to produce reports by category, country, and subject. These reports are available in hard copy and microfiche form and disseminated to all EPA libraries. The master file presently contains approximately 4,000 citations and abstracts.

(3) Document Control System

The Document Control System gathers data on the large collection of State and local environmental reports held by the many EPA libraries. The impetus for this system development was the fact that several of the larger libraries had individually developed automated systems for managing their own documents, and it was believed that a single system would provide standardization and cost savings without limiting the activities of the various libraries. The system actually consists of a single computer system and a number of separate files, one for each participating library. Microfiche copies of citations are provided.

In order to more effectively utilize the information contained on these files, the library is considering the creation of a consolidated master file which could acquaint each library with the information held in other libraries. This file would be used to generate microfiche citation files in author and title sequence. Title sequence would be utilized rather than subject sequence because of the variations which exist in assigning classifications in the various libraries. The ability of individual libraries to control their own document files would not be impaired.

(4) Air Ecology System

The Air Ecology System was developed by HEW at Research Triangle Park in North Carolina in cooperation with the

Department of Agriculture to gather bibliographic information on literature dealing with air pollutants and their effects on vegetation and animal life. The System was developed in addition to the existing APTIC system because it was felt that APTIC could not at that time satisfy the literature search needs of researchers in this field. After the establishment of EPA, the support staff for the system was transferred to the National Ecological Research Laboratory (NERL) in Corvallis, Oregon, where it presently supports a primary user group of approximately 30 researchers and technicians located in Corvallis; Raleigh, North Carolina and Research Triangle Park. While the Department of Agriculture also retained a copy of this system and its data base, the eventual disposition of that copy is currently unknown. The Air Ecology System maintains a reprint library of approximately 6800 articles covering the various effects of air pollutants on vegetation and animals. This collection contains much of the foreign literature produced, and approximately 300 of the non-English documents have been translated. This system is entirely maintained and operated under contract by the Oregon State University. The system generates microfiche citation listings and supports batch searching. Because of concerns regarding copywrite law, neither xerox nor microfiche copies of entire documents are provided.

Between 2400 and 2700 documents are being added annually to the data base. Selections are made by examining:

- Selective Dissemination Lists generated by commercial systems through the EPA Cincinnati Library, and funded by the EPA Library (approximately 30% of selections).

- Review of citation listings published by the Institute for Scientific Information (approximately 70% of selections).

- Suggestions and submissions by users.

Copies of documents selected are retrieved through requests to the author (50% successful), the Oregon State University Library, EPA Library or other interlibrary loan. Citations are then prepared and entered into the computer system through WYLBUR.

2. MANAGEMENT PLANS AND OBJECTIVES

Library Systems Branch Plans, Needs and Objectives related to the EPA Library Systems have been documented in the "PROGRAM MANAGEMENT OBJECTIVES REPORT" produced under this same Task Order. The primary factors discussed by management are noted below:

(1) Design Philosophy

Management believes that the Library Systems provide many of the benefits of more sophisticated systems at a much lower cost. It believes that this is primarily due to the design philosophy and considerations used in establishing file organizations, programming techniques, and input development. The basic approach used in developing each application was to establish a working system as soon as possible based upon a small test data base. This allowed each system to be developed, tested, modified, and implemented in a minimum amount of time and cost. Emphasis has been placed on applying various cost effective techniques in operating the systems such as maintaining master files on magnetic tape whenever possible and performing most file updates and report printing at night.

While Arthur Young and Company agrees that the systems are simple and cost effective on an individual basis, it does have several reservations regarding the approach used in designing a number of parallel systems to provide the same basic service, as discussed under 'Data Processing Characteristics'. A more carefully developed ADP Plan might have resulted in a smaller overall expenditure of cost and effort, and a more cohesive Library System.

(2) Working Tools for the Libraries

By making large amounts of information available to librarians and EPA personnel regardless of their location and by centralizing many of the clerical functions of the libraries, management feels these systems have reduced much of the librarian's workload burden. Thus the librarians are able to devote more time to directly serving users. In the eyes of management, the systems have upgraded the libraries and improved their staff from a professional point of view. Arthur Young and Company concurs with this approach and assessment.

(3) System Flexibility

Because the libraries function within a dynamic environment, systems should be adaptable to changing requirements. Management believes that its approach to systems development has led to the creation of flexible systems that can be rapidly and inexpensively modified to meet unanticipated needs. The approach being utilized provides a mixed blessing. While the systems apparently offer relative ease of modification, the state-of-the art in bibliographic systems is such that the vast majority of change should require no modification. It may be that the ease of implementing modifications encourages less than optimum initial design.

(4) Consolidation of Document Control and EPA Reports Files

Each library within EPA that wishes to utilize the centralized document control services for controlling its collection of state and local environmental reports is welcome to do so. Presently, each participating library maintains a separate master file for its own collection; however, plans are being laid to merge each file into one master file. When this task is completed, management further plans to combine the Document Control File and EPA Reports File. Because of similar record formats and processing flow, this task is not expected to require a great deal of effort.

(5) Transfer Funding Responsibility for Air Ecology System

Funding of the Air Ecology Computer System has been performed by the Library. However, in the next fiscal year, this responsibility will be transferred to the program group in Corvallis, Oregon. No plans have been made to modify this system since its primary users seem to feel it is quite adequate.

(6) Finding an Alternative for WYLBUR

Presently, the Library Systems utilize WYLBUR, a software package developed by Stanford University, for interactively updating and editing transaction files from remote terminals. Management feels that the computer costs associated with WYLBUR (approximately 10% of total data processing costs) are much too high and is presently seeking an alternative method for remote data entry. While not able to perform a detailed analysis on this subject, the project team has included some suggestions regarding this subject in Appendix F.

(7) Use of Mini-Computers

Management feels that perhaps long-term benefits and cost savings could be achieved by decentralizing much of the processing through the use of mini-computers. Many of the library systems could be processed less expensively and provide better service, management feels, by evolving in this manner. In addition, features of the mini-computer would support the development of an environmental thesaurus for cross-indexing subject terms between available bibliographic systems and provide for an automatic dial up capability to appropriate data bases in response to specific search criteria. While this subject could not be analyzed in detail during the project, the RECOMMENDATIONS AND ACTION PLAN and Appendix G address some of the inherent considerations.

3. ANALYSIS OF OBSERVED CHARACTERISTICS

This discussion delineates those characteristics of the EPA Library Systems (and Air Ecology System) which the project team considered to be most germane to the purposes of this study or in which particular improvements can be made. Where appropriate, comments are made regarding the impact of particular factors upon overall system performance.

(1) System Usage

Since the EPA Library Systems primary usage is on microfiche, and since statistics are not kept on librarian usage of the microfiche, it is impossible to identify the level and type of users. However, the documents referenced by the microfiche citations are an integral part of the library and, since the microfiche serve as the library catalog for these documents, it would be impossible to discontinue this product.

The Air Ecology System basically exists to serve the National Ecological Research Laboratory (NERL) in Corvallis, Oregon. The only other major user of this system is the United States Department of Agriculture, the agency which, along with HEW, originally developed the system. While microfiche copies of citations are provided to all regional libraries, some foreign governments and anyone else requesting it, it is impossible to measure the usage engendered, since they are performed manually by the end user.

The EPA Cincinnati Library performs a large number of searches on commercial data bases for users within the EPA. A summary of the searches and Selective Dissemination Activities provided by this library during 1975 is attached.

EXHIBIT I

EPA CINCINNATI LIBRARY
LIBRARY SEARCH/SELECTIVE DISSEMINATION ACTIVITY - 1975

	APTIC	BIO. ABSTS.	ENVIRON	FWRS	GEORGIA	IIT	RECON	LOCKHEED	MEDLINE*	NTIS	SDC	SSIE	SWIRS	TOXLINE	WRSIC	TOTALS
Reg. 1	2	4	0	0	13	0	4	30	1/4*	0	19	2	1	4	2	82/4*
Reg. 2	0	4	1	0	36	0	0	16	2	0	13	2	3	2	6	85
Reg. 3	3	2	4	0	56	0	0	31	8/10*	0	11	1	0	1	0	117/10*
Reg. 4	0	2	0	0	0	0	2	8	0	0	1	2	0	1	0	16
Reg. 5	0	1	0	0	0	0	10	50	4	0	21	4	0	21	9	120
Reg. 6	1	0	0	0	0	0	5	28	5/1*	0	7	2	0	10	2	60/1*
Reg. 7	0	0	2	0	70	0	1	17	0	0	1	1	0	3	4	99
Reg. 8	0	4	0	0	44	0	0	24	0	0	10	4	0	3	5	94
Reg. 9	2	0	0	0	57	0	0	8	1	0	10	1	0	2	0	81
Reg. 10	1	4	0	0	0	0	4	20	1	0	6	2	1	5	3	47
NERL CINTI	13	30	4	0	478	0	25	235	190/749*	0	100	14	2	96	10	1197/749*
NERL CORVALLIS	3	22	0	0	181	0	1	20	12/1*	0	15	5	0	1	4	264/1*
NERL LAS VEGAS	4	8	1	0	52	0	7	47	7	0	22	3	0	14	7	172
RTP	0	1	0	0	0	0	18	11	1/40*	0	0	0	0	0	4	35/40*
NWQLAB DULUTH	0	2	0	0	23	0	2	15	1	0	4	2	0	6	1	56
NMWQLAB NARAGANSET	0	14	0	3	120	0	2	14	10/51*	0	9	0	0	1	1	174/51*
RSKERR ADA, OKLA	0	1	0	0	0	0	6	13	5	0	9	0	1	1	1	37
GULF BREEZE, FLA	0	17	0	0	0	0	1	16	7	0	9	0	0	12	0	62
SERL LAB	0	0	0	0	0	0	0	10	1	0	4	0	0	5	0	20
MSAPC ANN ARBOR	5	0	0	0	0	0	0	36	0	0	19	3	0	1	0	64
EPA WASH	1	2	4	0	16	0	13	75	27/62*	0	61	27	2	32	10	270/62*
CDC	0	0	0	0	0	0	0	20	0	0	13	1	0	0	0	34
OTHERS	1	3	0	0	121	0	3	30	8/11*	0	15	4	1	4	3	193/11*
TOTALS	36	121	16	3	1267	0	104	774	291/929*	0	379	80	11	225	72	3379/929*

*Selective Dissemination Services, rather than retrospective searches

(2) Operations Costs

The annual costs for each of the EPA Library Systems include:

SERVICE	EXPENDITURE
EPA REPORTS	
Data Proccessing	$ 850
Salaries	1,081
NTIS (Report and Bibliography Publishing)	72,000
TOTAL EPA REPORTS	$73,931
INTERNATIONAL ENVIRONMENTAL REPORTING SYSTEM	
Data Processing (Input)	$ 1,344
Data Processing (Batch)	1,456
Salaries	1,081
Sigma Data	10,000
Library of Congress (Translation for Office of International Activities)	115,000
TOTAL INTERNATIONAL	$128,881
DOCUMENT CONTROL SYSTEM (HEADQUARTERS ONLY)	
Data Processing (Input)	$ 3,600
Data Processing (Batch)	1,800
Salaries	11,483
TOTAL DOCUMENT CONTROL	$16,883
SIGMA DATA	
(Contract for all services to the EPA Library, excluding International System)	
AIR ECOLOGY	
Data Processing (Input)	$ 2,880
Data Processing (Batch)	1,200
Oregon State University	40,997
TOTAL AIR ECOLOGY	$45,077

NOTE: These cost figures are based on cost breakdowns provided by the Library System Manager and the current Sigma Data and Oregon State University contracts. Most of the costs associated with NTIS (Preparation of Microfiche copies and bibliographies) and the Library of Congress (Translations and Citations) are not considered allocable to the Library Systems by the Project Team, since they are either required for acquisition of library materials (NTIS) or for the Office of International activities (Library of Congress translations).

In performing searches of external data bases, the EPA Cincinnati Library incurred the following costs not including salaries, during fiscal year 1974:

COMMERCIAL SEARCH COSTS (CINCINNATI)

EXTERNAL SYSTEM	EXPENDITURE
Smithsonian Institute	$ 4,800
Lockheed Laboratories	12,459
System Development Cooporation	11,909
Bioscience Information Services	13,063
University of Georgia	13,454
Medline	1,917
Toxline	2,461
Fish & Wildlife Service	360
TOTAL	$60,423

NOTE: With 4,308 searches performed during calendar year 1975, the overall cost per search, $14.03, corresponds fairly closely to the $15.00 used in our analyses of commercial data bases for others of the EPA Technical Literature Search Systems.

(3) Data Processing Characteristics

The data processing systems design employed by the EPA Library Systems and the Air Ecology System is a simple, effective library catalog design; it is not really a technical search system design. While these systems do provide a batch search capability, it is very infrequently used. This is fortunate, since a batch sequential search design is expensive and inefficient, as indicated in the SWIRS Data Processing Characteristics discussion.

Without a detailed understanding of the requests made of these systems, a judgement cannot be made regarding the responsiveness of this design to user needs. Normally, this type of system approach to searching will support more general or simple queries, addressing one or two specific topics, much more effectively than detailed queries which relate to several keywords simultaneously. However, it is not recommended that a formal technical literature search system be developed unless and until there is a proven, serious need for such a capability. If this need were to become apparent, it would exist across all the library holdings, not merely those reviewed during this study project.

(4) Source Overlap

Since the EPA Library Systems act more as automated catalog systems than technical literature search systems, the issue of source overlap is not critical. The only true candidate for substantial duplication is the EPA Reports System, whose contents are completely duplicated by the NTIS System. Previously the NTIS System did not allow the retrieval of citations on any element; the library requires retrieval on Contract Number/Source/Subject to identify those grants which have produced final reports and should be terminated. It is understood that NTIS now provides this capability, and can therefore be searched in lieu of a batch EPA Reports search. If the EPA Reports File is to be merged with the other EPA Library Files to produce a single microfiche catalog, this file may still be necessary. If not, its elimination should be considered.

It is probable that the Air Ecology System has substantial duplication with other data bases. The potential for duplication with the APTIC data base is high and should be researched. The fact that approximately 30% of the Air Ecology Data Base is selected through an examination of commercial data bases guarantees a duplication of at least 30% of all Air Ecology citations. Since most of the rest of the data base is selected by reviewing such commercially generated current awareness lists as those produced by the Institute for Scientific Information, it is probable that substantial duplication also exists in this area.

4. COMMERCIAL ABSTRACTING SERVICE

Commercial services do not appear to be a viable alternative to the EPA Library System. The Document Control System does not contain abstracts, and abstracts for the EPA Reports System are produced by the authors themselves. It is probable that some of the International Documents are being abstracted (and translated) commercially, and this possibility should be pursued. In terms of commercial data processing systems service, if the library is to remain in existence, it must have some mechanism for generating catalogs of its holdings.

Since no usage figures are available for the Air Ecology System (the OSU contractor has retained but never tabulated requests), no financial analysis can be made regarding reliance on commercial systems. In considering the sources of selections for the system, it certainly appears that commercial services would be an alternative. The analysis of this possibility could be economically and effectively performed in concert with that for the APTIC System.

5. PRIMARY CHARACTERISTICS

Arthur Young & Company has reservations concerning the methodology which has been used in the development of the EPA Library Systems. A single system design has been implemented over and over again in providing new systems for additional data bases. This development of parallel systems frequently appears economical on a temporary, day-to-day basis but, when examined over a longer period of time, is actually more expensive and less effective than the development of one system flexible enough to handle the several data bases involved. In fact, proper initial design should have resulted in a system and data base capable of supporting each of the three document types (and others) immediately. The planned merging of these files would have utilized no more than a computer sort/merge utility. Any new systems design effort might consider the development of a single, more flexible design which can process future bibliographic applications without modification.

Arthur Young & Company shares the Library System Manager's concern over the current total reliance of the Library upon the contractor, Sigma Data Computing Corporation. Contractors are most effectively utilized in performing discrete, defined tasks, which result in a completed data processing system or systems, trained users and operators, and complete program and system documentation. To do otherwise is to virtually ensure a system which is in a constant state of dependence, a lack of client expertise in their own system and insufficient or non-existent documentation. The current sole source contract under which the contractor is proceeding is indicative of the other adverse result: the 'locking in' of one particular company. The EPA Library should take steps to ensure the completion of all current tasks, with complete training and documentation, and perform all future developmental work through open bid contractors. While new or additional contractors would require some orientation to the system, the benefits of open competition would far outweigh that inconvenience.

The Air Ecology System is removed, both geographically and administratively, from the control of the EPA. Its necessity is unknown as well as its potential contributions to EPA activities other than the Corvallis National Ecological Research Laboratory. The individual to whom the Project Team was referred, and who must be considered the System Manager is an employee of the contractor, not of the EPA. The closest organizational ties of this system, in addition to the Corvallis National Ecological Research Laboratory, are with the Department of Agriculture. Since the Project Team was unable to visit this facility, it is recommended that the EPA identify the true usage of the system and determine its worth in supporting this one laboratory, particularly since the vast majority of its input is from secondary sources. The Air Ecology System was developed independently of the APTIC System because of a dissatisfaction with APTIC's turnaround time, quality of microfiche copies and indexing definition. While this independent development has avoided APTIC's current difficulties, the Project Team believes that if the Air Ecology System is truly beneficial, consideration should be given to consolidating it with APTIC. This should be done in a figurative sense at first, so that the steps recommended in this study can be performed considering a combined APTIC/Air Ecology System.

VII. RECOMMENDATIONS AND ACTION PLAN

Of the five EPA Technical Literature Systems analyzed during this project, the four major systems suffer from a single critical problem: a lack of effective, continuous managerial direction and control. The EPA Library Systems do not demonstrate the same degree of difficulty as the other systems, perhaps because of the more limited scope of the Library Systems, the fact that most of the Library Systems are created not from a selection and acquisition activity but from literature selected and provided by activities outside of the library, and their operation within a better defined and recognized framework.

Arthur Young & Company has located no EPA or Program Policy Statements regarding information dissemination and no written procedural statements, management objectives, performance requirements or system justifications at any management level, other than a few general system descriptions written by the System Managers themselves. Other than the EPA Orders regarding the library, only the most vague and general passages of various federal acts can be referenced by the systems for their support and direction.

The symptoms and products of this lack of management control have been apparent throughout this study project. Some have been referenced in our discussions of individual systems; others shall be discussed herein as a part of our RECOMMENDATIONS AND ACTION PLAN. Among the most serious results are:

- High cost in comparison to services rendered
- Very limited communication and coordination between individual systems, with other EPA program areas or with other Federal Agencies.
- Lack of effective communication and assistance from EPA service offices, such as the Office of Public Affairs and the Management Information and Data Systems Division.
- Continued development of myriad systems performing the same general activity.

It would be unjust to place the primary responsibility for system inadequacies upon the System Managers or even middle management. The System Managers have performed to the highest level which can be reasonably expected. They have been required by the EPA to be experts as librarians, public relations personnel, data systems analysts and financial managers. The lack of a definition of scope of these activities has caused them to attempt to provide a very wide range of services within a limited budget: Data Bases and Libraries containing all pertinent, non-duplicative literature, complete query and searching support without charge and, frequently, advertisement of the systems to all potential users. In brief, they are attempting more than their resources permit. Finally, it appears that the primary contact they have had with upper management or other EPA offices has been through continuing cutback orders without cutbacks in definitions of scope, and through attempts to eliminate their entire function. It is not surprising that the System Managers tend to isolate themselves and their function, or that they are not forthcoming with suggestions for reduction in scope or expenditures.

1. PROBLEM SYMPTOM AND RESULTS

The result of the lack of effective management has been an almost haphazard growth of the EPA Technical Literature Search Systems. Without the enforcement of proper research, planning and reporting, the systems have come to exhibit the following characteristics:

(1) Limited System Usage

None of the four Technical Information Systems enjoy a high level of usage, with the result that cost/usage figures are extremely high. (The Library systems do not retain data, since the microfiche bibliographies are utilized directly by the Regional Libraries). In addition, the characteristics of existing usage is not encouraging. A number of SWIRS requests do not warrant the time or specificity of a computer search, and neither SWIRS nor APTIC serves the population in which they are ostensibly most interested: the private citizen and citizens action groups. While users of the Pesticides System are in fact representatives of their chosen population, this population is not knowledgeable concerning the system. It seems only by accident that knowledge is gained of the system, while the vast majority of queries received are of such a general nature that a manual search would be as effective as a computer search. Part of the problem faced by the Pesticides System is due to the development in the Criteria and Evaluation (C&E) Division of a parallel capability. C&E employs its own search specialist, develops its own search strategies and frequently performs its own computer searches; many of the requests for very broad-scoped information are initiated by C&E.

The management of EPA should take immediate steps to determine the exact needs of present or potential users, the contribution which the EPA is prepared to make in responding to these needs, and the priorities which should be set in fulfilling them.

(2) Lack of Cost Data for Evaluation

Only the Library System Manager was able to provide reasonably accurate system cost data, and none appear to perform any cost-effectiveness analyses of the systems. Charges are not separated by individual manual or data processing functions. The Project Team experienced substantial difficulty in acquiring the cost data referenced in this report. Future contracts for outside services should contain the requirement for detailed recordation of costs and expenditures by functions, either on contractor invoices or as separate reports. System Managers should allocate the time of their own personnel according to function, and data processing services should be charged to differing account numbers which will identify the type of data collection and analysis so that System Managers can begin to improve their cost/benefit performance.

(3) Insufficient Knowledge of Usage and Holdings

Only the APTIC System can provide a breakdown of holdings in terms of subject, level of specificity and source. None of the systems retains information on items retrieved, so that none of the services can analyze the utilization of its data base. This capability is critical if System Management is to ensure the inclusion of literature most responsive to user needs. The most single and effective method is to add 'counter' fields to each record, to be used to accumulate the number of times that citation is selected, along with fields classifying the type and source of the document referenced. A relatively simple program can then be developed to provide an analysis of data usage.

(4) Lack of Data Processing Coordination and Expertise

Although only the new NOISE System Manager possesses appreciable experience and expertise in the field of data processing, each of the Systems Managers has independently developed and maintained the computer system supporting each of the Technical Information Systems. This approach ignores the benefits which should accrue from a reservoir of data processing expertise such as provided by the Management Information and Data Systems Division (MIDSD). The Project Team understands that there has been a liason position at MIDSD for technical literature systems, but does not consider this approach to be particularly effective, since the normal operations responsibilities of the MIDSD representaitve tends to overshadow his liason responsibilities.

The Project Team has also noted lack of communication between the Technical Information Services and MIDSD regarding dissatisfaction with commercial data processing services.

Arthur Young & Company has made no evaluation of any services provided by data processing vendors, and therefore cannot judge their performance. It is apparent, however, that several of the System Managers are dissatisfied with Optimum Systems, Inc. (OSI). Although MIDSD wishes to provide assistance to the System Managers in dealing with commercial vendors, the practice of providing notification to MIDSD only in the event of unresolveable difficulties precludes MIDSD from acquiring an understanding of the true situation. It is interesting to note that there has apparently been no response by MIDSD to written general criticisms of OSI's performance found in both the EPA Library's justification statement regarding a transfer of computer operations to the University of Cincinnati (produced during the summer of 1975) and the REGION X MINI-COMPUTER FEASIBILITY STUDY (produced on November 20, 1974).

Finally, the Project Team has noted that formal feasibility studies are not being prepared by the Technical Literature Services as required by the EPA Automatic Data Processing Manual, dated March 31, 1975. The Project Team has located only one document which might qualify as a feasibility study, the operational statement regarding transfer of the EPA Library computer operations to the University of Cincinnati. The NOISE examination of user needs, produced before the inception of the NOISE System, canot really be considered a feasibility study. The current computerization of the Pesticides Systems is not supported by a feasibility study, nor was the APTIC System development, or ongoing modifications and corrections. The ONAC Shared Information System is not supported by a written feasibility study. Arthur Young & Company has no reason to believe that such studies will be produced for any planned or potential enhancement or modifications to any of the systems.

(5) Lack of Coordination of Capabilities of the Several Systems

While the individual systems each contain some information which would be valuable to other systems, such as the EPA Library holdings of legislative and regulatory data, the Technical Literature Services do not utilize each other's holdings, either in collecting information or in satisfying search requests. While committees have been formed to provide cooperation at several times in the past, they have never provided an ongoing effective vehicle for communications and coordination.

The one concerted effort for system communications and coordination, which culminated in the December 10, 1973 SUMMARY OF JOINT MEETING OF APTIC, HAPAB AND SWIRS PROJECT OFFICERS WITH COMMON CONTRACTOR, apparently 'fizzled out' shortly thereafter. In most issues this report appears to be an explanation of why nothing could be done, rather than a blueprint for cooperation. In any event, the three systems have retained their unique processing streams and associated equipment, even though the intermediate and end-products are quite similar. It should be noted that although this document identifies only a slight overlap of journals whose review for pertinent documents is required by at least one of the three systems, the Project Team analysis (Appendix C) discovered a much higher degree of overlap.

(6) Lack of Cooperation with Other EPA Offices and Other Federal Agencies

Several potentials for interoffice and interagency cooperation exist which are not being realized. The Technology Transfer Program of the Office of Rsearch and Development (ORD) produces a number of 'How-to' Manuals, including such subjects as municipal waste processing and pollution control in selected industries, which would be particularly useful to a number of APTIC and SWIRS users. ORD also independently responds to a large number of requests for the public, and state, local and federal agencies. It appears that individuals with the same queries could easily have them answered by any of several different EPA entities with completely different materials. ORD is currently examining new methods of supplying information to the public in addition to or in concert with its Technology Transfer Publications and Seminars. It is wasteful for this activity to continue independently from the Technical Literature Systems and the Office of Public Affairs, which also provides environmental information to the public. The existence of a separate Information Office in the OPP Criteria and Evaluation Division poses the same type of problem, on an internal EPA basis, although the C&E Information Office recognizes the OPP Pesticides System and decides on an individual basis whether to utilize the OPP Technical Services Division. C&E does not currently have access to TOXLINE, which provides the only currently automated Pesticdes Data Base.

Finally, little effort has been expended by the EPA Technical Literature Services in pursuing cooperative agreements with other Federal and local agencies. There appears to be potential for cooperation between EPA and ERDA in energy-related pollution control subjects (primarily involving SWIRS and APTIC) and in other pollution-oriented activities of ERDA. The National Library of Medicine's TOXLINE System already contains the Pesticides Data Base and, while TOXLINE utilizes keywords, rather than the Pesticides Classification Codes, they would consider adding the Classification Codes. Arthur Young & Company is convinced that numerous other areas for cooperation can be identified with further research. As an example, a Washington newspaper recently carried an article discussing a new waste disposal bibliography activity being initiated by the Washington Suburban Sanitary Commission.

(7) Misallocation of Efforts

The EPA Technical Information Services appear to have established incorrect priorities in the allocation of funds and attention. The majority of effort expended by these services involves the maintenance of the data bases. Secondly, the services concentrate on responding to requests with the most comprehensive product possible. Only minimal effort is expended in publicizing the system and educating potential users. While this approach might be appropriate for a well-known, heavily-utilized system, it is the opposite of the requirements of these systems. The excellence of a library or the completeness of a data base is of little importance if a system is not used, and the current use of the EPA Technical Literature Systems is quite low.

Arthur Young & Company believes that, if these services are to fulfill their self-imposed missions, they must act to familiarize and educate their potential users. In the case of SWIRS and APTIC (and NOISE, if it wishes to aid the public), this requirement entails publicity of the service through all means available, including the media, government publications, seminars, environmental organizations and the rest of the EPA itself. As discussed under OPERATIONAL RECOMMENDATIONS, Arthur Young & Company believes that this change in emphasis can be accomplished without seriously degrading services provided.

The primary change in emphasis required of the Pesticides service involves the education of current OPP users. The prevalent habit of requesting everything about a particular compound is wasteful and unnecessary. Such a need can be filled by the requestor himself, simply by reviewing abstract bulletins. Requestors whose needs require detailed searches must be trained to identify their needs so that responsive searches can be made. Finally, it is imperative that the true information requirements of the upcoming registration activity be identified so that plans can be made to properly respond to them.

2. MANAGEMENT RECOMMENDATIONS

Arthur Young & Company recommends that each of the Technical Literature Activities be so structured, defined and operated as to provide a level of service appropriate to the organizational mission, existing level of need, and within available resources. The following recommendations delineate those steps which the Environmental Protection Agency should take in developing the management policies and organizational framework within which the Technical Information Activities can most effectively perform their assigned functions.

(1) EPA Must Establish a Comprehensive and Uniform Information Dissemination Policy

There is a critical need at the Environmental Protection Agency for an inclusive, Agency-wide policy regarding the dissemination of information regarding the causes and control of pollution. Arthur Young & Company does not believe that the individual Acts and Regulations referenced by the Technical Literature Services either mandate or preclude the type of information systems in operation at EPA. Rather, these systems have come into being either on the basis of decisions made at the middle management level or because of the assignment to EPA of other governmental units already supporting such systems. Consequently, before particular decisions can be made regarding the continued existence or characteristics of each of the systems, the missions and requirements of the service must be verified or redeveloped in concert with an overall policy on information dissemination. Since such policies do not currently exist, it is incumbent upon the EPA to examine its own missions in this light and develop this policy. This policy is also imperative for the determination of the proper working relationships between the EPA Technical Information Services, the EPA Library, the Office of Public Affairs, the Office of Research and Development and other EPA operating units.

While the development of an EPA Information Dissemination Policy is not within the scope of this project the following questions must be addressed in this developmment:

- Is the EPA to assemble and provide environmental information which was not developed by the EPA itself?
- Is this information to be available without charge, or is it to reflect the costs incurred in acquiring and disseminating it?

- If the literature is to be provided free of charge, will it be provided without charge to all requestors, or only to those who must implement the technical information contained?
- Is the general public to be encouraged to utilize this information? If so, is wide advertisement to be utilized (to encourage maximum participation), selective advertisement (to reach only those with a technical interest) or no advertisement (to ensure access only by those selected by EPA)?
- What differences exist in this policy for particular subject areas?
- To what degree are individual offices free to expand (or limit) this EPA policy as it relates to their own operations?
- Within each subject area, which office shall have the primary authority and responsibility to direct this information dissemination activity?
- What responsibility is to be assigned for determining user need and evaluating responsiveness to that need?

The current information dispersement policies of each of the systems appear to be as follows:

- SWIRS - Usage by anyone encouraged
- APTIC - Usage by anyone encouraged
- NOISE - Undecided; service in transition
- Pesticides - Usage by EPA only
- Library - Usage by EPA and other government

The Project Team's analysis is not dependent upon the chosen policy, since developed recommendations would be identical if usage of any of the systems were to triple. However, should a decision be made to stop supporting the public, including industry, the APTIC and SWIRS Systems would be imperiled, since they primarily support users located outside of the EPA.

(2) A Single, Central Technical Literature Coordinator is Needed to Plan and Monitor all Systems Activities.

It is recommended that an individual be assigned the position of Technical Literature Coordinator, to advise, coordinate and review the activities of each of the System Managers, in order to provide them with the expertise in financial management, program planning, user training and public relations activities which the systems currently are lacking. It is not recommended that the Technical Literature Services be transfered from their current organization position since they either support program needs or operate in tandem with other program activities. The Technical Literature Coordinator might occupy a staff position within the Office of Planning and Management, and would devote his full time to this position. His qualifications would include substantial ability and experience in financial management and control, statistical analysis and program planning and development. In this position he shall direct the individual System Managers in implementing the OPERATIONAL RECOMMENDATIONS and in correcting the PROBLEM SYMPTOMS AND RESULTS.

A Data Processing Specialist should be assigned to assist the Technical Literature Coordinator in any activities requiring computer expertise. This specialist should have experience in data base management systems and in input equipment selection. While the extent of his participation will depend upon the decisions made regarding the disposition of the systems, the Technical Literature Systems must be his primary responsibility, so that he can devote whatever time is necessary.

3. OPERATIONAL RECOMMENDATIONS

Within our discussions of each of the four primary Technical Information Systems (SWIRS, APTIC, NOISE and Pesticides), we have examined the Operations Costs of the systems in total and as a function of the number of requests and number of searches performed. The costs per search of these systems range from $234/search to $839/search. While it may be impossible to place a value upon timely, pertinent information, Arthur Young & Company considers these costs to be excessive, particularly since commercial services are available which may provide comparable information at substantially lower cost. As demonstrated in the COMMERCIAL ABSTRACTTING SERVICE analyses for each of the systems, savings from $188,000(Pesticides) to $516,700(APTIC) would result from a discontinuation of selection and acquisition activities and a reliance upon commercial services. These analyses also demonstrate that, even if it is necessary for the Technical Literature Services to continue selecting and abstracting at half their current level, donating the product to commercial services to ensure their inclusion, considerable savings could still result.

The primary advantages offered by commercial services include.

- Substantial decrease in operating cost
- Immediate improvement in SWIRS and APTIC service, since turnaround problems due to obsolete computer system design (SWIRS) and system problems and down time (APTIC) would be bypassed.

- Release of personnel time and funds to increase the recognition and sophistication of potential users, improve search quality, and perform the currently neglected management duties as discussed under PROBLEM SYMPTOMS AND RESULTS.

The potential disadvantages to this approach include:

- Specialized libraries would no longer be maintained, since they are currently maintained as a byproduct of the selection and acquisition process. While Arthur Young & Company is not able to judge the value of the specialized libraries, it is difficult to imagine a value equal to the savings which would result from a move to commercial services. The library might subscribe to those journals which are quite productive but very difficult to obtain and utilize the resources of other libraries for requests for articles in more common journals.

- EPA would be dependent upon the reliability of commercial systems. While this is true, the commercial systems have been continually upgrading their capability. With several capable commercial services currently in operation, this concern does not seem critical. While a problem in delay of data base updating was encountered by TOXLINE, it was not the type likely to repeat, and commercial services simply cannot afford to allow substantial delay. Finally, in most cases a delay due to down time is not critical, particularly when abstract bulletins are available for manual review.

- EPA would be dependent upon the pricing policy of the commercial service. It is impossible to predict the likelihood of substantial price increases, but with several major companies in the field there should be sufficient competition. Moreover, EPA, and the Federal Government overall, would be an extremely powerful customer and thus capable of ensuring reasonable prices.

- Commercially prepared abstracts may not be of the same quality or directed to the same users as those of the Technical Literature Systems.

While Arthur Young & Company is convinced of the economy of using commercial abstracting services and automated access systems, it is necessary to evaluate the completeness and usefulness of the commercial data bases. Each Technical Literature Service must verify the ability of the commercial data bases to properly respond to its own requirements. The following discussion lists, in chronological sequence, the steps which should be taken in evaluating commercial data services and, if necessary, improving or supplementing the products of these services.

(1) Review Commercial Data Base Coverage to Assure Search Adequacy

While overlap exists between the EPA Technical Literature Systems and the major commercial systems in terms of journals reviewed, it cannot be guaranteed that this same overlap exists in terms of articles selected. Therefore, for each of the systems being considered for replacement, a computer listing of the last 500 selections, sorted by author, should be generated. This listing could then be compared against a similar listing of the major commercial sources. The commercial services should be willing to generate such a list free of charge. If the vast majority of articles present on the EPA System is also present on the commercial system, EPA should begin to discontinue its own system and utilize the commercial system.

While individual System Managers have expressed the opinion that commercially prepared abstracts may not be of the same quality or written for the same audience as those of the EPA Systems, the Project Team has not found substantiation for this opinion. The review of commercial data bases should include an evaluation of the quality and usefulness of the commercial abstracts.

(2) Donation or Sale of EPA Data Bases to Commercial Services Would Insure Complete Coverage.

If no commercial data base(s) is discovered with a sufficient amount of duplicated selections with the corresponding EPA Data Base, EPA should attempt to sell or even donate its current data base to one or more of the commercial abstracting services. It is quite possible that one or more of the commercial abstracting services would agree to expand their coverage to include most EPA-related items, in return for the usage of the associated EPA data base. Even if such an agreement cannot be reached, it would be substantially cheaper for EPA to continue its selection and acquisition activity at 1/2 the current level, donate the results to a commercial service and use that service to search the overall data base.

This activity may present the EPA with a number of difficult choices. The Comparison Search Analysis (Appendix E), while not sufficiently inclusive to provide reliable evidence, indicates that commercial services may currently provide literature references similar but not identical to those provided by the EPA Technical Literature Search Services. If this should in fact be the case, each System Manager must examine the commercial data bases on a subject basis, through discussions with the commercial services and examinations of samples of the commercial data bases, to determine those subject areas which are adequately covered commercially. If the commercial service agrees to expand its coverage as discussed above, the EPA Technical Literature Service can still discontinue selection activity. Otherwise, EPA must determine, through further analysis of user queries and further interviews, those subjects which are critical to its present or anticipated user base and not adequately covered commercially. These would be the subjects in which EPA might continue its selection activity, for sale or donation to a commercial abstractor.

(3) If Commercial Service Data Bases are Unacceptable EPA Should Maintain a Single Technical Information Data Base

Only if each of the above approaches failed would Arthur Young & Company recommend the continuation of the Pesticide, SWIRS or APTIC data bases, and then under these conditions:

- EPA should purchase abstract tapes from commercial services rather than contracting the abstracting activity, for those journals and subject areas effectively covered by the commercial services.
- The resultant EPA data base should be run not on an EPA system but on one of the commercial services, such as SDC or Lockheed. These services are interested in purchasing the EPA data bases and servicing them just as a commercial data base. They would pay EPA a fixed fee and royalty, while EPA would pay a reduced rate to access its own data base.

(4) If EPA Operations of the Literature Search Systems is Essential a Redesign of Current Systems is Necessary

Should the EPA decide to retain its own data base for one or more of the Technical Literature Search Systems, it is not recommended that it utilize its own data processing system for any data base which it is able to put up on a commercial data processing system. However, if either SWIRS or APTIC is retained as a data processing system, then serious consideration should be given to designing a new data processing system providing on-line or batch searching using inverted files. This is recommended since both SWIRS and APTIC currently possess serious problems in terms of data processing performance and would cost more to upgrade than to redesign, particularly if an efficient data base management system is utilized. If the Pesticides System is retained, it could be left just as it is, as long as almost all searching is performed on microfiche. If frequent computerized searching should be begun for Pesticides, it should be redesigned along with SWIRS or APTIC.

If any EPA automated literature search capabilities are to be retained, they might best be supported through a single data processing system utilizing separate files for each system. This would allow individual services to retain control over their own files, while providing the benefits of a single data processing system and perhaps a common thesaurus. One method of developing such a system would be via a data base management system. Appendix G discusses the considerations to be taken in choosing a data base management system to provide for a bibliographic system. SYSTEM 2000, as it operates at RTP, would be a viable data base management system for such an application.

4. ADDITIONAL CONSIDERATIONS

Several additional factors have been addressed in this review of the EPA Technical Literature Search Systems.

(1) Minicomputers

There is an increasing interest in various operating units of the EPA in the use of minicomputers for 'specialized' applications. The determination of whether a minicomputer would provide an economical, effective approach must be made for each specific systems requirement. The current interest in minicomputers is not based on detailed analysis, but is rather a general interest in the concept of minicomputers. The cost analyses which have been performed to date are inadequate to support any decision to utilize or not utilize minicomputers.

The scope of this study did not include the detailed evaluation of minicomputer applications. However, Appendix G contains a general discussion of minicomputers and their applications. While there certainly exist applications for which minicomputers are appropriate, the following factors must be carefully considered:

- The desire for minicomputers is frequently based not on a technical need, but rather on a reaction to existing organizational problems, with a resulting desire to become independent from the central service
- Since a central data processing service does currently exist, the use of minicomputers, even when more economical for a particular operating division or office, can be more expensive for the total organization because of underutilization at the central facility

- The use of minicomputers frequently results in the development of independent data processing staffs, or contractors, within scattered divisions and offices. This situation results in distributing data processing and design capability over several areas with no central scheduling or personnel utilization control

- Minicomputers are not at the same level of sophistication as larger systems. At this time they provide more limited compilers and require greater programming expertise than larger systems, and are not as flexible. They do not support evolving applications to the same degree as larger computers.

- Minicomputers are only beginning to provide the type of flexible, simple generalized data base and retrieval system available for years on major computers. Consequently, users of higher level systems are performing as 'pioneers', and will encounter substantially more limitations and unresolved system problems than users of larger systems

- In part, because of the economy of minicomputer mainframes and the ease with which they can be expanded, it is not uncommon for data processing applications, and the minicomputer configuration supporting them, to grow to a midicomputer, with corresponding loss of any original economic benefits.

(2) Free Search Services

The EPA Technical Literature Systems are currently provided all services without charge. While the levying of a charge for services has been considered, it has been felt to be inconsistent with the mission and goals of the EPA. Arthur Young and Company recommends that users of the information systems be charged at least for the costs of retrieving information, for the following reasons:

- This is the only way in which the value of the systems may be judged. If users are cautious about spending $30 for a computer search, as a number of interviewees were, it is impossible to justify substantially higher expenditures as an information service

- The true cost of contracts and grants should include the cost of EPA-supplied information, even though this may be merely an interoffice transfer of funds

- To some slight degree, charges for services will provide needed funds for improvement of services and the further determination of user needs.

If the systems utilize commercial data processing system services or place their own data bases on commercial services, many industrial and academic users will go directly to the commercial system rather than the EPA and be charged through that mechanism. Then, if desired, users who do query EPA directly, probably including most private citizens and Citizen Action Groups, might still be serviced without charge.

(3) NOISE System Development

The Office of Noise Abatement and control has decided to create a single, integrated system which will act both as a bibliographic system and as an automated filing system of correspondence, developing reports, external documents and any other material considered valuable to overall ONAC operations. While not able to adjudge the need for this new NOISE system, Arthur Young and Company does recommend that any new NOISE development effort provide a system which can be utilized for future Agency-wide bibliographic applications. Since bibliographic systems requirements are basically standard, future permanent, temporary or specialized bibliographic systems should utilize this resulting system, and the proliferation of independent bibliographic data processing systems can be eliminated.

APPENDIX A

SAMPLE SEARCH ANALYSIS

During the EPA Technical Literature Systems Analysis, the Project Team reviewed a representative sample of actual requests of each of the four technical literature systems. These reviews were used as a primary basis for the identification of users, user needs and systems service, as described in the discussion of each systems. This appendix summarizes the information gathered for each system.

The following samples were taken:

- SWIRS; 500 searches for the period including November, 1974; May, 1975; August-October, 1975
- NOISE; 121 documents for the period June-October, 1975
- APTIC; 233 searches for the period June-August, 1975
- PESTICIDES: 191 documents for the period January-October, 1975 (statistics extracted from 166).

Explanation of terms: The following terms describe the headings of the charts in this Appendix.

- Level of Detail: degree of detail inherent in request

  D = Detailed
  S = Semi-detailed
  G = General

- Number of Computer Searches: actual number of computer searches
- Number Canned: SWIRS only; actual number of canned searches. A canned search is a previously run query; copy of the output is supplied instead of another computer search
- Average Turnaround: average number of days a request took to be processed
- Additional Reference List: a manually created list of bibliographic references supplied by the analyst which have not yet been abstracted.

- Number of Pages per Search: the number of Xeroxed pages sent in addition to the search material (copied from various catalogs).

- New User Packets: the number of new user packets sent to first-time users.

- W/O Notation: without notation; the number of searches processed but no data of the search recorded.

- No. HITS: The number of hits recorded during the computer search.

- W/O HITS: without hits: the number of computer searches that received no hits.

Following are exhibits depicting the statistics acquired during the Project Team's Analysis of each system. The level of record keeping by each system has impacted the statistics which could be developed.

UTILIZATION RESPONSE BY USER (EPA ONLY)
SWIRS SYSTEM

EXHIBIT II

USER	LEVEL OF DETAIL	NUMBER OF COMPUTER SEARCHES	NUMBER CANNED	AVERAGE TURNAROUND	ADDITIONAL REF. LIST	AVG. PAGES PER SEARCH	NEW USER PACKET
EPA WASH. D.C.	D	12	—	3.8	4	2.4	—
	S	5	—	11.0	4	6.4	—
	G	1	—	2.0	—	—	—
TOTAL		18		7.1			
EPA REGION I	D	—	—	—	—	—	—
	S	1	—	10.0	—	2.0	1
	G	—	—	—	—	—	—
TOTAL		1		10.0			
EPA REGION II	D	—	—	—	—	—	—
	S	1	—	9.0	—	—	1
	G	1	—	1.0	—	—	—
TOTAL		2		5.0			
EPA REGION III	D	1	—	10.0	—	—	1
	S	—	—	—	—	—	—
	G	—	—	—	—	—	—
TOTAL		1		10.0			
EPA REGION IV	D	1	—	5.0	—	—	—
	S	—	—	—	—	—	—
	G	—	—	—	—	—	—
TOTAL		1		5.0			
EPA REGION V	D	5	1	8.0	1	8.0	—
	S	—	—	—	—	—	—
	G	—	—	—	—	—	—
TOTAL		5		8.0			
EPA REGION VII	D	5	1	7.2	3	11.5	—
	S	6	—	14.0	6	1.1	1
	G	—	—	—	—	—	—
TOTAL		11		11.5			
EPA REGION VIII	D	3	—	7.7	1	0.3	—
	S	—	—	—	—	—	—
	G	—	—	—	—	—	—
TOTAL		3		7.7			
EPA REGION IX	D	1	—	3.0	—	—	—
	S	—	—	—	—	—	—
	G	—	—	—	—	—	—
TOTAL		1		3.0			
EPA Region X	D	1	—	8.0	—	—	—
	S	—	—	—	1	4.0	—
	G	—	—	—		—	—
TOTAL		1		8.0			
TOTALS		44	2	8.1	20	4.1	4

Preceding page blank

UTILIZATION RESPONSE BY USER (excluding EPA)
SWIRS SYSTEM

EXHIBIT III

SUBJECT AREA	LEVEL OF DETAIL	NUMBER OF COMPUTER SEARCHES	NUMBER CANNED	AVERAGE TURNAROUND	ADDITIONAL REFERENCE LIST	AVERAGE PAGES PER SEARCH	NEW USER PACKETS
U.S. GOV'T excluding EPA	D	11	—	17.5	4	1.4	1
	S	9	3	6.6	—	3.3	3
	G	1	—	16.0	2	15.0	—
TOTAL		21	3	12.0			4
STATE/ LOCAL GOVERNMENT	D	19	—	8.6	4	4.6	4
	S	36	2	9.5	11	4.2	8
	G	—	5	8.2	—	—	3
TOTAL		55	7	9.1			15
FOREIGN GOVERNMENT	D	11	—	7.3	[illegible]	2.6	2
	S	5	—	17.4		—	—
	G	1	—	65.0		—	—
TOTAL		17		13.6			2
INDUSTRIAL/ COMMERCIAL	D	68	1	10.4	13	2.8	18
	S	81	4	12.2	11	2.9	11
	G	1	3	2.8	-	—	2
TOTAL		150	8	11.2			31
CIVIC ORGANIZATION	D	2	—	12.5	1	12.0	—
	S	6	1	16.1	4	9.[illegible]	3
	G	—	1	2.0	—	—	—
TOTAL		8	2	14.0			3
COLLEGE PROFESSORS	D	26	1	5.0	9	3.1	6
	S	34	3	10.5	8	3.7	2
	G	—	7	4.4	—	10.4	4
TOTAL		60	11	7.8			12
COLLEGE STUDENTS	D	9	—	8.7	3	3.2	3
	S	14	—	19.6	3	7.0	3
	G	—	2	3.5	—	—	1
TOTAL		23	2	14.3			7
PRIVATE CITIZENS	D	18	—	16.7	4	4.8	5
	S	34	6	12.5	15	3.8	10
	G	1	28	4.4	1	4.1	16
TOTAL		53	34	10.7			31
TOTALS		387	67	10.7	96	3.6	105

SWIRS

EXHIBIT IV. NUMBER OF SEARCHES BY GROUP

Group	Number	Percentage
Industrial/Commercial	158	31.6
Private Citizens	87	17.4
College Professors	71	14.2
State/Local Government	62	12.4
EPA	46	9.2
College Students	25	4.8
U.S. Government (excluding EPA)	24	4.8
Foreign Government	17	3.4
Civic Organizations	10	2.0
	500	100.0

Industrial users utilize the system almost twice as much as the next nearest group.

SWIRS

EXHIBIT V. AVERAGE TURNAROUND FOR SEARCHES

Group Category	Turnaround	Ranking
College Professors	7.8	1
EPA Only	8.1	2
State/Local Government	9.1	3
Private Citizens	10.7	4
Industrial/Commercial	11.2	5
U.S. Government (excluding EPA)	12.0	6
Foreign Government	13.6	7
Civic Organizations	14.0	8
College Students	14.3	9

SWIRS

EXHIBIT XV. SOURCE OF KNOWLEDGE OF SWIRS AS INDICATED BY REQUESTS

Mother Earth News

Environment Magazine

Psychology Today

Update - a Solid Waste Publication

Unnamed Magazine

Educational Source

U. S. Gov't. Agency referral

EPA Referral

Franklin Institute Referral

NOISE

EXHIBIT VII. USER UTILIZATION (EPA ONLY)

		NOISE SYSTEM NUMBER OF SEARCHES					SHARED ACQUISITIONS LIST NUMBER OF SEARCHES				
USER	LEVEL OF DETAIL	ALL	W/O HITS	W/O NOTATION	NO. HITS	Avg.	All	W/O HITS	W/O NOTATION	NO. HITS	AVG
STANDARDS	D	16	—	1	741	49.4	59	24	—	38	1.1
&	S	4	1	—	67	22.3	—	—	—	—	—
REGULATIONS	G	8	—	5	613	204.3	4	—	2	7	3.5
TOTAL		28	1	6	1421		63	24	2	45	
TECHNICAL	D	5	—	1	5	1.3	7	—	2	6	1.2
ASSISTANCE &	S	4	—	2	35	17.5	2	—	1	1	1.0
OPERATIONS	G	2	—	2	—	0	—	—	—	—	—
TOTAL		11	0	5	40		9	0	3	7	
PROGRAM	D	—	—	—	—	—	9	6	1	2	
PLANNING	S	—	—	—	—	—	—	—	—	—	—
	G	—	—	—	—	—	—	—	—	—	—
TOTAL		0	0	0	0		9	6	1	2	
TOTAL		39	1	11	1461	54.1	81	30	6	54	1.2

NOISE

EXHIBIT VIII. USER UTILIZATION (EXCLUDING EPA)

		NOISE SYSTEM					SHARED ACQUISITIONS LIST				
		NUMBER OF SEARCHES					NUMBER OF SEARCHES				
USER	LEVEL OF DETAIL	ALL	W/O HITS	W/O NOTATION	NO. HITS	AVG.	ALL	W/O HITS	W/O NOTATION	NO. HITS	AVG.
INFORMATICS	D	—	—	—	—	—	6	1	—	5	1.0
	S	—	—	—	—	—	—	—	—	—	—
	G	—	—	—	—	—	—	—	—	—	—
INDUSTRIAL/ COMMERICAL	D	—	—	—	—	—	2	—	—	1	0.5
	S	1	—	1	—	0	—	—	—	—	—
	G	—	—	—	—	—	—	—	—	—	—
U.S. GOV'T (EXCLUDING EPA)	D	1	—	1	—	0	—	—	—	—	—
	S	—	—	—	—	—	—	—	—	—	—
	G	—	—	—	—	—	—	—	—	—	—
FOREIGN GOV'TS	D	—	—	—	—	—	1	—	1	—	0
	S	—	—	—	—	—	—	—	—	—	—
	G	—	—	—	—	—	—	—	—	—	—
TOTAL		2	—	2	—	0	9	1	1	6	0.9

APTIC

EXHIBIT IX. USER UTILIZATION (EPA ONLY)

	LEVEL OF DETAIL	NUMBER OF COMPUTER SEARCHES	NUMBER OF REQUESTS	AVG. NO. OF SEARCHES PER REQUEST	AVG. TURNAROUND PER SEARCH
REGION I	D	—	—	—	—
	S	2	1	2.0	0.5
	G	—	—	—	—
TOTAL		2	1		
REGION II	D	—	—	—	—
	S	—	—	—	—
	G	1	1	1.0	6.0
TOTAL		1	1		
REGION III	D	1	1	1.0	12.0
	S	1	1	1.0	1.0
	G	—	—	—	—
TOTAL		2	2		
REGION IV	D	9	6	1.5	20.7
	S	9	8	1.0	29.7
	G	6	6	1.0	22.3
TOTAL		24	20		
REGION V	D	1	1	1.0	27.0
	S	1	1	1.0	28.0
	G	1	1	1.0	20.0
TOTAL		3	3		
REGION X	D	1	1	1.0	29.0
	S	—	—	—	—
	G	—	—	—	—
TOTAL		1	1		
Total		33	28		21.5

Preceding page blank

APTIC

EXHIBIT X. USER UTILIZATION (EXCLUDING EPA)

USER	LEVEL OF DETAIL	NUMBER OF COMPUTER SEARCHES	NUMBER OF REQUESTS	AVG NO. OF SEARCHES PER REQUEST	AVG. TURNAROUND PER SEARCH
U.S. GOV'T (EXCLUDING EPA)	D	1	1	1.0	27.0
	S	11	5	2.2	21.4
	G				
TOTAL		12	6		
STATE/LOCAL GOV'T	D	23	15	2.1	27.8
	S	11	10	1.1	21.0
	G	1	1	1.0	5.0
TOTAL		35	26		
FOREIGN GOV'T	D	4	5	0.8	60.3
	S	--	--	--	--
	G	1	2	1.0	35.0
TOTAL		5	7		
INDUSTRIAL/ COMMERCIAL	D	33	19	1.7	30.2
	S	54	31	1.7	24.0
	G	12	10	1.2	24.6
TOTAL		99	60		
COLLEGE PROFESSORS	D	20	10	2.0	32.5
	S	10	7	1.4	35.0
	G	6	3	2.0	13.7
TOTAL		36	20		
COLLEGE STUDENTS	D	1	1	1.0	56.0
	S	--	--	--	--
	G	2	2	1.0	38.0
TOTAL		3	3		

USER UTILIZATION (EXCLUDING EPA)

USER	LEVEL OF DETAIL	NUMBER OF COMPUTER SEARCHES	NUMBER OF REQUESTS	AVG. NO. OF SEARCHES PER REQUEST	AVG. TURNAROUND PER SEARCH
PRIVATE CITIZENS	D	2	2	1.0	25.5
	S	7	3	2.3	61.3
	G	1	1	1.0	25.0
TOTAL		10	6		
TOTALS		200	126	1.6	28.6

APTIC

EXHIBIT XI. UTILIZATION BY SUBJECT (EPA ONLY)

SUBJECT	LEVEL OF DETAIL	NUMBER OF COMPUTER SEARCHES	NUMBER OF REQUESTS	AVG NO. OF SEARCHES PER REQUEST	AVG. TURNAROUND PER SEARCH
MEASUREMENT METHODS	D	11	7	1.8	18.9
	S	1	1	1.0	1.0
	G	--	--	--	--
TOTAL		12	8		
EMISSION SOURCES	D	--	--	--	--
	S	1	1	1.0	27.0
	G	--	--	--	--
TOTAL		1	1		
EFFECTS ON PLANTS & LIVESTOCK	D	1	1	1.0	28.0
	S	2	2	1.0	32.0
	G	--	--	--	--
TOTAL		3	3		
CONTROL METHODS	D	2	2	1.0	29.5
	S	2	1	2.0	0.5
	G	1	1	1.0	26.0
TOTAL		5	4		
BASIC SCIENCE & TECHNOLOGY	D	1	1	1.0	42.0
	S	--	--	--	--
	G	--	--	--	--
TOTAL		1	1		
MISCELLANEOUS	D	--	--	--	--
	S	--	--	--	--
	G	9	9	1.0	18.5
TOTAL		9	9		
EFFECT ON HUMAN HEALTH	D	--	--	--	--
	S	2	2	1.0	7.5
	G	--	--	--	--
TOTAL		2	2		
TOTALS		33	28	1.2	19.3

APTIC

EXHIBIT XII. UTILIZATION BY SUBJECT (EXCLUDING EPA)

SUBJECT	LEVEL OF DETAIL	NUMBER OF COMPUTER SEARCHES	NUMBER OF REQUESTS	AVG NO. OF SEARCHES PER REQUEST	AVG. TURNAROUND PER SEARCH
MEASUREMENT METHODS	D	10	6	1.7	31.3
	S	5	5	1.0	10.0
	G	3	1	3.0	19.0
TOTAL		18	12		
LEGAL & ADMINISTRATIVE	D	--	--	--	--
	S	2	2	1.0	30.0
	G	--	--	--	--
TOTAL		2	2		
EMISSION SOURCE	D	2	2	1.0	21.0
	S	5	3	1.7	25.0
	G	2	2	1.0	35.0
TOTAL		9	7		
EFFECTS ON PLANTS & LIVESTOCK	D	10	6	1.7	28.5
	S	8	5	1.6	46.2
	G	--	--	--	--
TOTAL		18	11		
MULTIPLE CATEGORIES	D	6	5	1.2	24.0
	S	6	3	2.0	13.5
	G	--	--	--	--
TOTAL		12	8		
CONTROL METHODS	D	29	18	1.6	30.3
	S	19	10	1.9	30.2
	G	2	2	1.0	36.0
TOTAL		50	30		
ATMOSPHERIC INTERACTION	D	3	2	1.5	29.0
	S	5	3	1.7	24.0
	G	2	2	1.0	35.0
TOTAL		10	7		

APTIC

UTILIZATION BY SUBJECT (EXCLUDING EPA)

SUBJECT	LEVEL OF DETAIL	NUMBER OF COMPUTER SEARCHES	NUMBER OF REQUESTS	AVG NO. OF SEARCHES PER REQUEST	TURNAROUND PER SEARCH
BASIC SCIENCE	D	13	7	1.9	33.9
&	S	12	6	2.0	24.0
TECHNOLOGY	G	--	-	-	
TOTAL		25	13		
MISCELLANEOUS	D	2	2	1.0	21.0
	S	15	10	1.5	27.6
	G	12	7	1.7	13.6
TOTAL		29	19		
AIR QUALITY	D	-	-	---	----
MEASUREMENT	S	5	3	1.7	16.3
	G	-	-	---	----
TOTAL		5	3		
EFFECTS ON	D	12	6	4.0	10.5
HUMAN HEALTH	S	5	5	1.0	33.0
	G	5	3	1.5	24.3
TOTAL		22	14		
TOTALS		200	126	1.6	25.8

PESTICIDES

EXHIBIT XIII. AVERAGE HITS BY DATA SEARCHED

SYSTEM	NO. HITS	AVG. HITS PER SEARCH
TOXLINE	3204	97.1
MEDLINE	918	35.3
CANCERLINE	195	12.2
SDILINE	34	3.8
EXCERPTA MEDICA	573	35.8
CHEM ABSTRACTS	14	14.0
CAIN	8	8.0
CHEMLINE	50	16.7

NOTE: These figures only include hits; zero hits are not averaged in. Absence of complete documentation did not permit a more definitive chart.

COMPOUND/CONCEPT REQUESTS

The following table can be used to indicate the level of detail contained in individual requests. Requests for compounds only are requests which asked for all available information on a particular compound. Concepts identify effects of compounds or other subject areas regarding pesticides. Consequently, the linking of compound and concepts narrows the request to a specific problem area.

Type of Request	Number	Percentage
Requests for Compounds only	128	77.1%
Requests for Concepts only	3	1.8%
Requests for Compounds and Concepts	35	21.1%
Total Sample	166	100%

Average Turnaround per request: 10.6 days (approximately 20% of the requests examined did not have completed dates; these were not included in the statistics).

Preceding page blank

APPENDIX B
USER SURVEYS

In order to develop an understanding of user attitudes, needs and evaluations of the various systems, the Project Team interviewed a number of users selected from the search requests analyzed in the SAMPLE SEARCH ANALYSIS. This appendix contains the results of these interviews, and contains a sample of the interview form utilized.

The primary subject areas addressed in the interview included:

- Source of user knowledge of the system
- Reason the information was desired
- User sophistication regarding the service and knowledge of other sources
- Actual use of the information
- Service quality
- Importance of the service to the user
- Suggestions

All interviews were conducted by telephone and took approximately 20-30 minutes each. The number and type of interviewees is indicated in the beginning of each system discussion.

1. SWIRS

(1) Interviews

Eleven users from EPA and two commercial users were interviewed. All were aware of the SWIRS system although five people have only used the system once. The other eight averaged past usage of 11 times. There were several sources of knowledge of SWIRS: three people found out through general office knowledge; four learned of it from a presentation by EPA personnel in Cincinnati; three called EPA in Cincinnati and were informed of the system; two learned of it from a booklet (one identified the booklet as "Finding Your Way Through EPA" dated October, 1973); one person learned of the service through a presentation by a group from the Small Business Administration.

Information from the system was used for a variety of reasons: two people used the service to provide background information for themselves and for contractor support. One person used the service to assist an outside agency. Three people used the service only for background information, because of a new assignment or to become aware of the state-of-the-art. Five people needed the information for a project; one was doing a research paper and one just wanted to test the data base for an example of its content.

Five interviewees were well aware of other sources and bibliographic services and had, in fact, used them. One person had explored Chemical Abstracts and the Library of Congress but said he did not know of any commercial systems. The other seven were not at all aware of any commercial systems. Twelve people used the information received for the original purpose. One received no information but believed this an important indication of what is not available.

Quality of service provided a diverse response. The following matrix depicts this:

Quality Point	Yes	No	No Opinion
Timely *	13 *	-	-
Understandable	12	-	1
Pertinent	11 **	-	2
Too much information	3	9	1
Too little information	4	7	2
Poor quality (illegible)	4	8	1

*Although interviewees stated they were satisfied with the timeliness it is important to be aware of the time frame involved. Charts in Appendix A identify average turnaround time for EPA users as 8.1 days. However, note that this does not include mailing delays (i.e., request mailed from user to Cincinnati or to SWIRS, and the time lag for return). This total period often involves two to three weeks.

**One person questioned the currency of the data.

In response to the question regarding the contribution the information made, six people felt SWIRS provided a significant contribution. One person did not know and two felt the question did not apply to their original need (i.e. background or information only). One stated that the service helped but was not critical to the project and two considered the service marginally helpful. One person felt the information for his particular need was too old, therefore, not required.

Eleven respondents stated that they could get copies of the articles (or felt they could) without much difficulty. Two did not try because they were not needed.

As far as repeat usage of the system, eleven stated they would use the system again. One would have to be somewhat assured of good response prior to using it again. It would be dependent upon the nature of the query. The other stated that he would use any means available to get information but felt that a technical researcher would do a better job. However, he acknowledged that the service does save time. Seven interviewees stated they would not alter usage of the system if a charge per search of $30.00 was initiated and one of the six felt such a charge would be counter-productive. Five others felt they would use more discretion with searches (e.g., a bit more caution) and one person felt he would not use the system unless he was sure to get some good answers.

(2) Criticisms and Suggestions

Lastly, all interviewees were asked for criticisms and suggestions.

Criticisms

Articles are all pie-in-the-sky speeches; not enough performance; too much on what was going to be done.

Did not have the required forms in his office for query. OSWMP personnel had not sent them out on time.

SWIRS not interactive with other systems, should be on-line.

Difficult to identify contents of the data base.

System not direct enough to the user nor fast enough.

Not sure of input criteria.

Abstracting is too vague; needs to be more meaningful; more meat; less off-the-cuff remarks.

Abstracts too wide to paste on 3 x 5 cards or print too large.

System does not cover anything in the field.

Abstracts too lengthy.

Unsure how thoroughly the literature was covered and updated.

Suggestions

Publicize the system; how to use it; what it contains; source of abstracts; currency; how it works.

Speed up the timeliness.

Should be able to go straight to the data base rather than have to go through Cincinnati.

Eliminate Demonstration Grants announcements.

Would prefer completed project results.

Eliminate the Job Control language and other "computer talk" on the output.

Improve paper quality (microfilm copy).

Make use of the system mandatory as it saves time and money.

Use a greater range of sources for information.

2. NOISE

(1) Interviews

Nine people from EPA were contacted but four could offer no response as they had little contact with the system. Trying to make contact with people in spite of changed offices, new telephone numbers, or personnel being away from their office, presented great difficulty and consumed large amounts of time. Knowledge of the NOISE system was general office knowledge and all were multiple users of at least one query per month. The purpose of the queries were for project support, background information, and support for other people. Most people use other sources of information but all interviewees were not aware of commercial systems.

According to the respondents, quality of the service was satisfactory.

Quality Point	Yes	No	No Opinion
Timeliness	4	1	-
Understandable	4	1	-
Pertinent	4	-	1
Too Much Information	1	4	-
Too Little Information	-	4	1
Poor Quality (Illegible)	-	4	1

One respondent stated that turnaround took two weeks or more before he received abstract copies. No one expressed any difficulty acquiring completed articles. Most people felt that their work could be completed without the service, but with some difficulty. At either extreme, one said completion of the work would be impossible while the other said it would not have made any difference without the service. However, all respondents would continue to use the service. Only one responded in a positive manner to eliminate continued use if a charge were levied.

(2) Criticisms and Suggestions

Lastly, all respondents were asked for criticisms and suggestions.

Criticisms

Users are not aware of the contents of the data base.

Data base is not as comprehensive as it should be.

Data base needs more technical information in it.

The literature is too repetitive.

Suggestions

Change the format - highlight salient points and provide tables and charts.

Tie into other systems.

Increase definition or detail - data base is not comprehensive enough.

3. APTIC

(1) Interviews

Nine persons were interviewed: four from EPA, three commercial users, one other Federal Agency, and one from academia.

Only four interviewees were knowledgeable of commercial systems. Two people heard of APTIC from an EPA booklet while the other users knew of the system or heard about it from a friend. All except four wanted the information for use on projects. Three submitted the request for other users and one wanted information for his dissertation. All but three used the information for its intended purpose. The three were not end-users and could not provide input to this question. All, except those who were end-users, said the service contributed significantly to their work and all except one felt they would use the system again. The one could not state whether or not he would reuse the system.

The following table depicts user response to the question of quality of service.

Quality Point	Yes	No	No Opinion
Timeliness *	5	4	-
Understandable	8	-	1
Pertinent	7	-	2
Too much information	2	5	2
Too little information	1	5	3
Poor Quality (illegible)	-	7	2

* Although five people were satisfied with the timeliness one recognized that there have been turnaround problems. Further, one person stated that it took over a month to receive abstracts. The APTIC Utilization Charts in Appendix A will show that turnaround varies from 24.0 days to 42.0 days (not including mail and possible administrative delays both to and from APTIC).

All users, except one who was not an end-user, stated it would not be difficult to get complete articles. When asked if use of the service would continue if there were a $30.00 charge, six people said yes, two did not know, and one was doubtful.

(2) Criticisms and Suggestions

Lastly, respondents were asked for criticisms and suggestions.

Criticisms

It is criminal not to have the system under Lockheed or S.D.C.

APTIC does not pick up books and conferences.

APTIC has outlived its usefulness.

Too much overlap with other systems (MEDLINE, BIOSIS, CHEM ABSTRACTS).

Suggestions

Improve turnaround

Refine the thesaurus as its vocabulary is too limited.

Improve turnaround.

4. PESTICIDES

(1) Interviews

Ten persons from EPA were interviewed. All except one were interested in the service for support of an EPA related project or study. One person wanted the information for background use. Only two persons were aware of a separate EPA system called PESTICIDES. All except one were knowledgeable to varying degrees of commercial systems. Knowledge of the service came from the respondents office in most cases, while two were informed by EPA Library personnel and one learned of the system through the National Institutes of Health. All except one agreed that the impact of the service and information was significant. Several stated that their projects could not be completed without this service.

Users most often requested searches of very broad scope (containing no more than a compound name) because a narrow search was too exclusive; they were afraid something would be missed. One stated that he would narrow it down if the topic demanded it. Others divided up the various parts of the output to members of their project team.

The following table depicts user response to the question of quality of service.

Quality Point	Yes	No	No Opinion
Timeliness*	6	2	2
Understandable	10	-	-
Pertinent	8	1	1
Too much information	2	8	-
Too little information	3	5	2
Poor quality (illegible)	1**	9	-

* Although most users stated they were satisfied with the timeliness, it is important to note that turnaround time has encompassed a period of up to a month-and-one-half.

** Interviewee stated that abstracting was poor.

Regardless of the quality of the system all interviewees felt they would continue to use the system. However, if a charge of $30.00 were to be levied for each search, eight would continue to use the service, one was undecided and one would not unless policy were set by management. This same person feels he could perform a better search manually. The rest would exercise more caution in their use.

A significant problem was acquisition of the complete article. One person was satisfied with the acquisition process, four have no idea of any difficulty involved but have not tried to get copies. One person, who had not used this aspect, was aware of slow turnaround. Four people were quite emphatic in stating their difficulty in getting complete articles. Estimates for receipt of articles were one week to one year through the EPA library. If fast turnaround is required, George Washington University is used. However, a charge is levied for this service.

(2) Criticisms and Suggestions

Lastly, respondents were asked for criticisms and suggestions.

Criticisms

The data base only goes back to 1970 so it cannot be relied on.

Keywords are too exclusive. The system needs to be on-line so it can be "paged" through.

Two persons stated that they have to depend on others to write the strategy and do the search.

Suggestions

Two people stated the scope of the system should be publicized. Three mentioned expanding the number of years maintained on the data base.

One person stated the system should be able to search all derivatives of a compound rather than the user having to specify each and every name.

Quicker turnaround.

Get hardcopy faster.

INTERVIEW FORMS

TECHNICAL LITERATURE SEARCH SYSTEMS

EXHIBIT XIV

LITERATURE SYSTEM ____________________

SEARCH NUMBER ____________________

USER NAME ____________________

ORGANIZATION ____________________

ADDRESS ____________________

____

TELEPHONE NO. ____________________

1. How did you learn of the bibliographic service?

____

____

____

2. How many times have you used the service?

____

____

____

3. For what purpose did you submit this request? (What did you want to do with the information?)

____

____

____

4. Did you use any other sources of information (other services, library, etc.,)?

____

____

____

5. Why did you use this service rather than a commercial system (TOXLINE, NTIS, etc.,) or the EPA library or public library?

6. What was the quality of the service?

   Timely ____________________
   Understandable ____________________
   Pertinent ____________________
   Too much information ____________________
   Too little information ____________________
   Poor quality (illegible) ____________________

7. Were you able to get copies of the articles recommended by the service? Where (EPA library, public library, came with abstracts, etc.,)?

8. Did you actually use the information? Did you use the articles mentioned? How?

9. How difficult would it have been to complete your project without the EPA-supplied information? (Did the information contribute significantly to your project?)

10. Will you use this service the next time you have a similar need? Why or why not?

11. If a charge were levied for the user of the service, would you continue to use it?

$30 ____________________

$20 ____________________

$10 ____________________

12. Do you have any suggestions for improvement of the service?

____

____

____

APPENDIX C

ANALYSIS OF REQUIRED JOURNALS

As a means of determining the potential for duplication of coverage in each of the EPA Technical Literature Systems, whether among themselves or with existing commercial abstracting services, samples of the required journal lists of each system were compared to each other and to the lists of journals reviewed by the commercial services. A required journal is one which the particular Technical Literature Service requires their contractor to review in performing the selection activity. Commerical services compared were those which were most likely to review journals similar to those reviewed by the particular Technical Literature System.

1. OVERLAP WITH COMMERCIAL SERVICES

The following tables identify, for each Technical Literature System, the number of journals sampled and the number and percentage of this sample reviewed by various commercial services examined.

SWIRS SYSTEM	NUMBER OF JOURNALS	PERCENTAGE
TOTAL SAMPLE	153	100%
Reviewed by at least one of the below	109	71.2%
Chemical Abstracts (CA)	95	62.1%
Engineering Index (EI)	38	24.8%
Both CA and EI	19	12.4%

NOTE : Sample included journals beginning with A,B. or H.

NOISE SYSTEM	NUMBER OF JOURNALS	PERCENTAGE
TOTAL SAMPLE	146	100%
Reviewed by at least one of the below	97	66.4%
Chemical Abstracts (CA)	76	52.7%
Engineering Index (EI)	29	19.9%
Biological Abstracts (BA)	47	32.2%
Both CA and EI	17	11.6%
Both CA and BA	37	25.3%
Both BA and EI	3	2.1%
CA and EI and BA	2	1.4%

NOTE: Sample included journals beginning with A,H. or O.

APTIC	NUMBER OF JOURNALS	PERCENTAGE
TOTAL SAMPLE	298	100%
Reviewed by at least one of the below	271	90.0%
Chemical Abstract (CA)	244	81.9%
Engineering Index (EI)	55	18.5%
Index Medicas (IM)	86	28.9%
Both CA and EI	51	17.1%
Both CA and IM	85	28.5%
Both EI and IM	3	1.0%
CA and EI and IM	3	1.0%

NOTE: Samples included journals beginning with A,C, or N.

PESTICIDES SYSTEM	NUMBER OF JOURNALS	PERCENTAGE
TOTAL SAMPLE	281	100%
Reviewed by at least one of the below	230	81.9%
Chemical Abstracts (CA)	198	70.5%
Index Medicas (IM)	92	32.7%
Biological Abstracts (BA)	163	58.0%
Both CA and IM	80	28.5%
Both CA and BA	133	47.3%
Both IM and BA	76	27.0%
CA and BA and IM	70	24.9%

NOTE: Samples included journals beginning with A,H or O.

2. OVERLAP AMONG SYSTEMS

In comparing required journals of each Technical Literature System to each of the other systems, the sample consisted of all journals whose title began with A,B,C,D,H,L,O,U, X or Y.

On a percentage basis, the overlap among the systems is displayed below:

SYSTEM	SWIRS	NOISE	APTIC	PESTICIDES
SWIRS		11.0%	36.7%	15.1%
NOISE	12.6%	-	24.7%	13.4%
APTIC	35.3%	21.1%	-	37.1%
PESTICIDES	25.0%	7.7%	9.9%	-

Because the number of required journals for each system differs, the number of journals in the sample also differs:

SYSTEM	NUMBER OF JOURNALS
SWIRS	272
NOISE	239
APTIC	289
PESTICIDES	416

The following table displays the amount of overlap among various groups of system journals:

SYSTEMS	NUMBER OF COMMON TITLES
SWIRS/APTIC	99
SWIRS/PESTICIDES	41
SWIRS/NOISE	30
APTIC/PESTICIDES	104
APTIC/NOISE	59
PESTICIDES/NOISE	32
SWIRS/PESTICIDES/APTIC	23
APTIC/SWIRS/NOISE	12
APTIC/NOISE/PESTICIDES	10
SWIRS/NOISE/PESTICIDES	3
SWIRS/NOISE/APTIC	3
SWIRS/NOISE/APTIC/PESTICIDES	3

APPENDIX D

APTIC DATA PROCESSING PROBLEM LOG

The following list of problems encountered by APTIC are taken from the APTIC Problem Log maintained by APTIC at Research Triangle Park. The log depicts, by date, problems encountered during the processing of the APTIC System.

DATE	PROBLEM
1975	
January 6	Wrong Master File name caused job to abort.
January 8	Job aborted due to "bad" tape, processing storage requirements were exceeded and job was "lost" in computer.
February 4	Storage requirements exceeded: job aborted
February 5	Storage requirements exceeded: job aborted
February 6	Storage requirements exceeded: job aborted
February 12	Storage requirements exceeded: job aborted
February 12	Job resubmitted using card input: ran successfully Note: This job was originally submitted on January 8.
February 13	Storage requirements exceeded: job aborted
February 21	Storage requirements exceeded: job aborted
February 24	File dictionary "lost": cannot submit jobs to process without it.
February 25	File dictionary "lost": cannot submit jobs to process without it.
February 26	File dictionary "lost": cannot submit jobs to process without it.

NOTE: The log was not maintained from January 9 - February 3 since the person who maintains the log was on sick leave.

DATE	PROBLEM
1975	
February 27	File dictionary "lost": cannot submit jobs to process without it.
March 7	Storage requirements exceeded: job aborted
March 10	Citation file "gone"
March 11	Storage requirements exceeded: job aborted
	A note at this point identified the inability to Edit because of a program not working
March 18	Equipment check. Job to be resubmitted
March 20	Batch key failure
March 25	Job failed
March 26	On backlog
March 27	On backlog
March 28	Program "lost"
March 31	Jobs still have not processed
April 1	Jobs still have not processed
April 2	"Bad" tape replaced from March 27 job
April 4	Received print from March 27 job
April 7	No Inverted file
April 8	No Inverted file
April 9	Exceeded maximum time: aborted
April 10	Exceeded maximum time: aborted
April 11	Exceeded maximum time: aborted
April 15	Exceeded maximum time: aborted
April 18	Exceeded maximum time: one of two jobs aborted.

DATES 1975	PROBLEM
April 22	First program in backlog; received wrong print for second program.
April 23	Received printout for first program
April 25	Second job in backlog
April 28	Received printout for second job
April 30	Job aborted: many jobs "stopped by Univac" which were supposed to be run.
April 30	All searches submitted April 24 were "lost"
May 9	Submitted six searches; only received one back
May 20	Submitted 33 searches; only received one back
May 21	Searches from May 20 found.
May 22	Job stopped while in backlog; Received printout only for last search submitted.
May 27	Received printout for first program submitted on May 23. Second program printout expected later in day. Part of first printout bad due to printer job submitted
May 29	Job submitted on May 28 in backlog
May 30	In backlog (from May 28)
June 2	In backlog (from May 28)
June 3	Received printout but no searches.
June 6	Only received last search.
June 10	Jobs submitted on June 9 in backlog
June 11	Jobs submitted on June 9 in backlog
June 12	Jobs submitted on June 9 in backlog
June 13	Jobs submitted on June 9 in backlog

DATE	PROBLEM
1975	
June 16	Jobs submitted on June 9 not processed; disk system down.
June 17	Jobs aborted (submitted on June 9) due to sort error. Job separated and then ran.
June 23	Two jobs submitted on June 20. One failed due to lack of sort space. When failed job separated into two jobs, first one failed and second was in backlog.
June 24	Second job still in backlog; second half of the other job failed due to lack of sort space. Resubmitted.
June 25	In backlog
June 26	In backlog
June 27	Job failed; resubmitted
June 28	Job failed; resubmitted
July 1	Two of three jobs failed; resubmitted
July 2	Jobs in backlog
July 3	Jobs in backlog
July 5	Jobs failed; not enough disk space available.
July 7	Jobs in backlog
July 8	One job processed; others in backlog
July 9	Jobs failed; resubmitted
July 10	Jobs failed
July 11	One job ran, other failed.
	Note: Log was not available for period July 12-November 3.
November 4	Deck set up incorrectly; aborted four jobs
November 11	Storage requirements exceeded; print did not finish

DATE 1975	PROBLEM
November 18	Bad tape label, Job aborted Job exceeded maximum time, Job on backlog cancelled.
November 19	Restoring computer cancelled two jobs
November 20	Storage requirements exceeded; job on backlog cancelled.
November 21	Job exceeded maximum time; storage requirement exceeded.
November 24	Storage requirements exceeded; file dropped from Master Directory aborted three jobs
November 26	Job exceeded maximum time; parity error on Master file tape
December 1	Difficulty in reading Master File tape
December 2	Hardware error reading tape, four jobs cancelled
December 5	Job aborted due to unspecified processing step; two jobs exceeded maximum time; one job "disappeared"
December 8	Two jobs from December 5 ran OK; two cancelled because file was dropped from Master Directory.
December 12	Two jobs exceeded run time.
December 15	Five jobs "lost" due to system error
December 17	Five jobs cancelled by operator due to system failure
December 18	One job failed due to keypunch error; one job exceeded maximum run time; two jobs cancelled by operator
December 22	One job failed due to card missing from deck; one job exceeded maximum run time; two jobs in backlog
December 23	Five jobs failed because End of File card missing.

APPENDIX E
COMPARISON SEARCHES

As a means of estimating the proportion of the EPA Technical Literature Systems data bases which are also present on major commercial systems, the project team selected ten sample searches for each system to be performed on the major commercial bases. These searches were performed by the Information Retrieval Libraries at the EPA Cincinnati Library, using whatever data bases that researcher deemed most appropriate. Time and funding constraints dictated that only a few of the searches be run. The Pesticides comparison was performed in a different manner. Since Pesticides normally searches commercial data bases, the Project Team simply acquired several complete Pesticides searches and compared entries from Pesticides (found on TOXLINE) and those from other data bases. The extent to which this could be done was limited, since Pesticides runs such general searches that search output is voluminous. The search reviewed represents one of the smaller outputs.

The results of this experiment have been inconclusive. In some cases the commercial data bases provided for more hits on a subject than the EPA System, and in others the opposite is true. Very few duplicates were present. Some of the reasons for these differences may be:

- Differences in indexing methods. Several of the EPA Thesauri were much more limited than those of commercial systems, in part because the EPA systems include a more limited overall scope. Difficulty was experienced by the Information Retrieval Librarian in determining exactly what information was indicated by the group

- Since this was a first-time experiment, it was difficult to choose the data bases which would be most productive, or most similar to the EPA data bases.

Arthur Young and Company does not believe that these results can support a decision to eliminate (or retain) the EPA selection activity and data bases. The steps which should be taken to determine the viability of commercial data bases as replacements for the EPA Systems are identified in the RECOMMENDATIONS AND ACTION PLAN.

Following are the questions used for comparison of commercial sources and the EPA Technical Literature Systems.

- SWIRS

  # 6) Site selection for a regional Solid Waste Management Program

  # 9) Monsanto Landgard pyrolysis process being used in Baltimore, Maryland from January 1974 to present

- APTIC

  # 2) Information relating to air pollution effects on cotton

  # 9) Information regarding environmental pollution originating from acrylamide, urea-formaldehyde, urethan, when these materials are used for soil consolidation, soil coagulation, soil stabilization, soil conditioning

  #11) Information relating to the effects of air pollution on lipid metabolism

- NOISE

  # 4) Information regarding physiological effects, auditory hazards, damage risk criteria of quarry noise, mining noise, construction noise

  # 8) Information relating to vegetation attenuation

- PESTICIDES

  Dichlorophenoxgeocetate, -cetic acid (2,4-D)

The following table compares the 'hits' found on the original Technical Literature System search to those found on the commercial data base. The number of EPA citations returned which were also returned by one or more of the commercial systems are found in the 'No. of Duplicates" column.

EXHIBIT XV

NUMBER OF HITS BY DATA BASE

SYSTEM/ QUERY NO.	EPA SYSTEM HITS	CA	NTIS	EI	MEDLINE	BA	CAIN	POL	EM	NO. OF DUPLICATES
SWIRS #6	30	59	78	32	N/A	N/A	N/A	N/A	N/A	2
SWIRS #9	14	0	1	2	N/A	N/A	N/A	7	N/A	1/2*
APTIC #2	196	3	1	N/A	N/A	4	2	3	N/A	5/7*
APTIC #9	4	57	4	1	N/A	N/A	N/A	N/A	N/A	0
APTIC #11	177	N/A	N/A	N/A	23	N/A	N/A	N/A	N/A	8
NOISE #4	25	N/A	12	N/A	109	N/A	N/A	N/A	N/A	2
NOISE #8	43	0	0	N/A	N/A	6	19	11	N/A	10/16*
PESTICIDES	62	429	22	N/A	370	884	N/A	N/A	59	15/27**

*First number indicates number of citations matched by one commercial system; Second number indicates total number of citations duplicated by one or more commercial systems; Difference between first and second numbers indicates articles present on two or more commercial systems.

**Only 27 Pesticide hits were examined for duplicates; 15 were found.

APPENDIX F

ALTERNATIVES TO WYLBUR

As an integral part of costs the system used to input and output the data should be examined. EPA is utilizing a time sharing system called WYLBUR for this purpose. In order to properly address the subject of time sharing, it is appropriate to discuss some advantages and disadvantages of this process.

There are several benefits to using remote computing facilities (i.e., time sharing). Flexibility in use with languages and editing; interaction capability between man and machine; quick turnaround; user application programs for specific needs; networking (which provides scattered users access to centralized data bases); and dedicated services (the operation of your terminal as if the computer were dedicated to your use only).

Conversely, disadvantages are often overlooked. They include input slowed to the speed of the input/output terminal (speeds of 1100 lines per minute will not be realized in a 30 line per minute terminal printer); low efficiency of computational programs due to need for large amounts of central processor time and memory space. As a result of this, some system's efficiency is low (this, normally, is not a problem with bibliographic systems); safeguarding material can be difficult if security provisions of the system are not adequate; "saturation" can slow processing considerally (when too many users access the system at the same time); high communications and high data storage costs; and training personnel on the use of the system can be time consuming.

Of special importance is EPA's use of WYLBUR as a data input and output facility. It must be noted here that WYLBUR is a program which is a text editor and remote job entry facility. It is not a data entry or data output facility. Utilization as such is an expensive misuse of WYLBUR. Several EPA systems use WYLBUR for job entry; (i.e., SWIRS, PESTICIDES, and the LIBRARY SYSTEMS). While WYLBUR can be used in this manner, it is not cost effective. There are connect time costs, and resource costs which are computed in the overall charge for computer services.

Diverse pricing policies make cost comparisons difficult because the costs are computed in various ways. Some companies charge a single all-inclusive fee, some offer specialized billing on a per-transaction or per-item basis. Most bill for central processer usage and terminal connect time. Some charge for mass storage, initiation of service, and special applications program usage. Due to the myriad pricing schemes no standard rate can be established hence no cost analysis can be provided here.

DATA ENTRY

There are several methods of data entry which can be compared and should be mentioned here including; Punched cards, magnetic tape, disk packs, and magnetic card/selectric typewriter.

- Punched cards are a known and familiar feature, easily read (when interpreted) easily handled, and can be sorted and merged prior to computer input. Cost per thousand characters of storage is about 7.3 cents. Verification of punched cards is a necessary part of the punched card data input process.

- Key to tape and key to disk (or diskette) data entry methods are similar types of entry of data but differ in terms of storage medium and flexibility of machine. While retaining the characteristics of a keypunch machine, processing is significantly improved in terms of speed, program changes, verification, large amounts of storage, and visual display of data with the attachment of a cathode ray tube (CRT). A CRT lets the operator see what has been "punched" and make necessary editing changes. Key to tape and key to disk may be used with a high speed (2000 Baud or more) communications interface to transmit data to the host computer.

- Magnetic tape can only be read by machine. Because of the ease in erasing and reusing magnetic tape, care must be taken in its handling and storage. Cost per 1000 characters of storage at 1600 bits per inch is about .035 cents, but specialized equipment is needed to handle this type of input.

- Magnetic disk packs can only be read by machine. As with magnetic tape, disk packs can easily be erased and reused. Therefore, care in handling and storage is extremely important. There are large disk packs which hold about 25 million characters on 20 recording surfaces to a diskette (one surface, about the size of a 45 RPM record) which can hold about 93,000 characters. Costs per 1000 characters of storage range from about 4.96 cents for a large disk pack to about .92 cents for a diskette. As with magnetic tape, specialized devices are needed which are quite costly.

- The magnetic card/selectric typewriter (MC/ST) is relatively new as a data input device. In this case a secretary types the data onto a small magnetic card which fits into a machine attached to a selectric typewriter. Data on the finished card may be stored for later use or immediately input into the computer data bases through the MC/ST communicating feature over telephone lines. Rental costs for one MC/ST station are about $255.00/month. Purchase costs are about $9,550.00 (Amortized over five years would be about $1,910.00/year.) In quantity, magnetic cards would cost about $.90 each and they are reusable. Backup support for typing in lieu of abstracting would also be provided through this means.

- An adjunct to keypunch equipment is the "buffered" keypunch. This provides for quiet operation, changes and corrections to be made to data in the buffer before the card is punched, verification, and the changes to program formats. Similar in operation to a keypunch it offers speed and flexibility as key characteristics.

RECOMMENDATIONS

For an abstracting and editing operation with high volume, the best overall cost effective equipment would consist either of a combination of key to tape, disk or diskette with a CRT, or a MC/ST with communicating features. A multi-station key to tape/disk operation or an MC/ST is cost-effective in a high volume environment, when several formats are needed, when procedural editing can eliminate formal verification (i.e., column for column as in an 80 column card), and when a professional staff of abstractors can keep the equipment operators in a high production environment.

The cost effectiveness of these systems can be realized providing full utilization of the equipment. Instead of using WYLBUR as an input medium, input can be entered directly onto a tape which can be entered into the computer. This is a cost effective use of the system and a proper use of WYLBUR. The costs involved here range from about $6,000.00 for a very basic station to a minicomputer equivalent of $30,000.00. MC/ST's and key/disk or key/tape input devices are thus similar in cost with each offering their own unique capabilities.

APPENDIX G

DATA BASE SYSTEMS AND MINICOMPUTERS

In reviewing bibliographic systems, consideration was given to alternative methods of supporting them as a replacement for current processing methods. Two such alternatives are discussed here in the form of a single data base management system supporting commonly, all bibliographic systems, and the use of minicomputers resident in the EPA Program area to support each application uniquely. Advantages and disadvantages of these alternatives are discussed primarily from the technical standpoint with few value judgements made as to the feasibility of each in supporting individual systems. Rather, it is our interest here to provide an overview for subsequent evaluation of each as a viable alternative to current processing procedure.

It is appropriate to establish a framework for this discussion by describing necessary and desirable attributes of a bibliographic system.

I. CHARACTERISTICS OF A BIBLIOGRAPHIC SYSTEM

Literature search systems consist of a search capability either in an on-line or batch mode, edit update functions, and a bibliographic or abstract journal reporting program. On-line support offers the capability for searching an author, title, subject, etc. as provided by the applications programs.

The material or contents of the data base may be structured in a linear fashion or in an inverted manner. Contents generally include author, title, publisher, date, abstract, and key word vocabulary. Additional content such as subject categories, geographical locations, etc. are often included.

Data entry begins with preparation of the article through abstracting and indexing services of a skilled technical information specialist. Actual entry to the system may be via batch, key entry, or scanning. Output usually takes the form of management interest and user interest reports. Data on costs, time, number of queries, etc. are for management accounting. User reports take the form of computer printouts of the bibliography or abstract. The output can also be in the form of microfilm or microfiche from which the bibliography or abstract may be copied.

A. Types of Bibliographic Searches

Bibliographic search systems are currently able to perform three types of searches:

- Retrospective
- Demand
- Selective Dissemination of Information.

A retrospective search enables one to query previously published literature concerning a particular subject matter. Often times the result of a retrospective search will be a printed catalogue of references of a given subject. A demand search is one which provides for responses to a given detailed query against a data base, and is frequently used when a researcher is beginning a new project or wishes to review existing literature on a limited subject. Selective dissemination of information is more of a personalized service where the system notifies the users of the existence of current literature which may be of value to them based on a pre-established individual interest profile expressed in the thesaurus terms of the bibliographic system. The difference between selective dissemination of information and a retrospective search is that the queries represent interest profiles of the users consisting of several subject areas rather than one particular subject area.

Recently a new development has opened a fourth type of capability: time sharing. The introduction of time sharing has added a new dimension to bibliographic retrieval. Individuals may now interrogate data bases in an on-line interactive mode. In this manner the user is not required to formulate his entire plan in advance. Instead, he can browse, requesting assistance and repeatedly modifying his search strategy. This exchange can continue until the user has retrieved a sufficient number of abstracts to satisfy his needs. A number of on-line bibliographic systems have been developed which provide for such a dialogue to take place between the user and the system. These dialogues must, of course, ensure an effective instructional facility. These facilities help the user learn how to interact with the system.

B. System Features

In order to be able to select those features which are appropriate to a bibliographic system, each should be examined and its utility weighed against its cost. Both required and desirable features are summarized below.

Of key importance in all systems are the features that should be included:

- Instructional and diagnostic features
- On-line input features
- On-line output features

Instructional and diagnostic features should include the following:

- An easily understood user guide oriented towards non-programming personnel
- On-line documentation which offers brief descriptions of all commands, interrogatives and error messages
- On-line training
- A section for data base overview which will contain information concerning the size of the data base, the type of material in it, the time span covered, informational fields, descriptions of the various records in the files, examples of special search arguments, sample searches and even a pocket card with the command names and quick reference information.

On-line input features are designed to address the very real problems of structuring queries in such a way as to maximize value to the user.

- A request set which can be specified such that the results of earlier queries can easily be incorporated into later ones
- Logical operators such as OR, AND, and NOT help to broaden a query, narrow a query, or exclude terms
- A dictionary which will permit a display of short alphabetical portions of the index. This aids in queueing the user to terms which, in alphabetical sequence, will relate to terms used in his query which do not exactly match to the index
- A search field control would permit the user to indicate specific items or fields of a record which are to be searched
- Suffix removal would be specified by the user using the root or stem of a word followed by a truncation code. Very broadly speaking, such an example might be using Sodium as opposed to having to write out Sodium Chloride.

- Relational operators such as EQUAL TO, GREATER THAN, LESS THAN
- Spelling variations such as February, and Feb.
- Hierarchical thesaurus word proximity operators which allow the user to specify that index terms must all fall within the same field in order to qualify
- Phrase decomposition where natural language phrasing is automatically decomposed into significant words ignoring common words like the, a, and, etc.
- A facility for search profiles where the user can develop his own search strategy, store that strategy, re-run it at a later date after he up-dates it
- Sequential searching as contrasted to index searching would permit the system to scan fields which have not been previously indexed.

On-line output features would include:

- Sophisticated report generation to provide the capability to direct photocopy printing
- Search review which would provide a short summary of request sets which are currently active
- An on-line output formatting capability which would let the user at the terminal control the fields he wishes displayed such as placement of data on the output page.
- An off-line printing capability which would allow voluminous data to be printed on a high speed printer
- A pre-defined output format to provide the user with optional ways of viewing retrieved results prior to their being printed
- A rapid scan feature to allow the user to scan as much or as little of the retrieved information as desired
- Highlighting which permits one to know which specific terms in the abstract caused it to be matched with the key words. (Upper and lower case letter capability should be provided)

- Expanding, which provides for displaying the complete record after seeing a portion of it

- A sorting capability which would output records ordered according to the value of designated fields

- A ranking which lists documents in the order of decreasing number of search terms appearing in each record

- Special interface to an off-line photo composition or printing system such as computer output microfilm

- Data access protection which restricts access to the data base or to various levels within the data base only to users who know the required passwords

- A special terminal which provides for display of stored microfiche of a retrieved entry

- The capability to display graphs

- Statistical interface which would permit the retrieved data to be passed to an on-line statistical package

- A computing capability to allow such transactions to take place

- An audit trail of specified data such as number of "hits," citation number, etc.

The foregoing is not an idealistic approach. These facilities are available. However, each would have a cost, and optional features as indicated above may not always be necessary.

C. Required Features of a Bibliographic System

All of the previously described features, while "nice to have," are not necessary and would be costly if adopted totally. Therefore, some capabilities are considered more as luxuries and need not be part of a fully operational bibliographic system. Those features considered necessary are summarized below.

- A complete users guide and system documentation

- If on-line: Instructional and diagnostic features to interact with the user

- If on-line: Request sets (results of earlier queries incorporated into later ones)

- Logical and Relational Operators

- Search Field Control

- If on-line: Dictionary Access

- If on-line: Search Review, output formatting, and predefined output formats

- Off-line Printing.

II. DATA BASE MANAGEMENT SYSTEMS

A data management system is a software system which permits access to, and retrieval of, files. It performs functions such as information retrieval, report generation and individual inquiry applications. A data base management system is a software system which, in addition to the above, maintains data in a specified structure to enable multiple applications. Data is organized and retains relationships between different data elements and within data elements of the data base.

Exhibit XXV compares several Data Base Management Systems in terms of major bibliographic considerations. System 2000 is capable of supporting a bibliographic application, and can be used as an example of the data base concept. A portion of the Introduction to System 2000 Reference Manual provides a brief summary of major system capabilities as utilized on the UNIVAC computer:

- SYSTEM 2000 is a general-purpose data base management system which operates on IBM 360 and 370 series, UNIVAC 1100 series and CDC 6000 and CYBER series computers. The Basic SYSTEM 2000, with selected optional features, provides the basis for developing information systems tailored to the requirements of the application and the user. The following provides a brief overview of the basic system and each of the optional features.

BASIC SYSTEM 2000

- Basic SYSTEM 2000 provides the user with a comprehensive set of data base management capabilities. These include the ability to define new data bases, modify the definition of existing data bases, and to retrieve and update values in these data bases.

- The basic components of data base definitions are data elements and repeating groups. Values are stored in data elements. Repeating groups describe a structure for storing multiple sets of data values and also serve to link hierarchical levels of the definition.

- Values for each element and logical entry (record) may vary in length. The user may specify without restriction which elements in the data base are to be inverted and become key fields, and what hierarchical relationship an element will have with other elements in the data base. Data security is maintained by password control to the data base and additional password control to each component.

- Basic SYSTEM 2000 provides archival copies of data bases and records an audit trail of changes made to a data base. It is capable of reconstructing a data base by applying the audit trail, completely or in part, to an archival copy of that data base.

PROCEDURAL LANGUAGE FEATURE

- The Procedural Language Feature enables users to manipulate data in a SYSTEM 2000 data base from a procedural programming language such as COBOL, FORTRAN, or assembly language. This feature provides the mechanism for addressing any part of the data base of interest to the procedural program, to retrieve data in a sequence and format suitable for procedural processing, and to update the data base from the program. Interrelationships between two or more data bases can be established which permit network data structure to be defined.

- Data base qualification is performed by use of an index file, facilitating the very rapid screening of data of interest. Retrieved data may be sorted on one or more keys prior to return of the first set of data to the program. A complete set of status flags is provided to the programs to permit easy checking of abnormal or terminal conditions.

IMMEDIATE ACCESS FEATURE

- The Immediate Access Feature provides a user-oriented language with which a non-programmer may express his requests for retrieval or updating of a data base. The English-like language is easily learned. It includes a complete set of easy-to-understand diagnostic messages and is highly suited for interactive use from remote keyboard terminals.

REPORT WRITER FEATURE

- The Report Writer Feature enables the user to prepare report definitions in which he may generate breakpoints on any data base element or repeating group; specify headings and footings for each physical page, logical page, and the report as a whole; and control printing, item inclusion, page ejection, and the accumulation of sub- or grand totals dynamically via conditional statements. Up to 100 reports can be generated from a single pass of the data base index files.

SEQUENTIAL FILE FEATURE

For UNIVAC installations, the Sequential File Feature allows the SYSTEM 2000 user to create and access data bases residing on magnetic tape. Either the Report Writer feature or a Procedural Language program may be used to access a sequential data base. A link capability within the Procedural Language feature permits the user to establish logical associations between physically separate data bases. A sequential data base may thus be viewed as a logical extension of a direct access data base.

Without the benefit of simulation testing, it is impossible to determine the efficiency of the System 2000 in terms of update processing. However, if file structure and design is carefully prepared, the System 2000 update costs should be only moderately higher than those of a special purpose bibliographic systems using inverted files. Considering the small amounts of data normally retrieved through a bibliographic system on a single pass, any well-designed inverted file system will be more efficient than a sequential system. The primary disadvantages of System 2000 are that it does not support retrieval using only a portion of a particular keyword, does not automatically search the thesaurus for spelling variations and does not support full text searching These disadvantages, as compared to a specialized bibliographic system, must be weighed against the facts that the system is implemented at EPA, expertise exists at the EPA and conversion would be relatively straightforward.

DATA BASE MANAGEMENT SYSTEMS COMPARISON

	ADABAS	IMS	SYSTEM 2000	TOTAL	IDMS	DL1
Numeric Oriented Text Oriented	Yes No	Yes No	Yes No	Yes No	Yes No	Yes No
Retrieval Oriented Update Oriented	Yes No	No Yes	Yes No	No Yes	No Yes	No Yes
Ad Hoc Query	Yes	No	Yes	No	No	No
Security Controls	Yes	Yes	Yes	Uses Host Language	Privacy Lock	Yes
Computer System	IBM 360/370	IBM 370	IBM 360/370/ UNIVAC	IBM 360/370	IBM 360/370	IBM 370
System Accounting Information	Yes	Yes	Yes	No	With Add'l. Program (CICS)	With Add'l. Program (CICS)
Multi-File Searching	No	No	No	No	No	No
COBOL, FORTRAN PL1, ASSEMBLER	Not FORTRAN	Yes	Yes	Yes + RPGII	Yes	Not FORTRAN
Supports Inverted Structure	Yes	Yes	Yes	Yes	Yes	Yes

EXHIBIT XVI

III. MINICOMPUTERS

The advent of the minicomputer has brought a small revolution to the application of digital computing power to the solution of problems previously not computerized. The term "minicomputer" is somewhat of a misnomer as, generally, the minicomputers available today have similar capabilities as their larger counterparts. "Mini" relates more to price and portability than to capability although often, anticipated "mini" costs can expand well beyond expectation.

The minicomputer, as generally applied to MIS or business problems, offers the user his own computing environment which he can use under his control for his own applications exclusively. For the purposes of this discussion a minicomputer may have the following characteristics:

- Stand-alone configuration including memory, central processing unit, input/output channels and devices operating under stored program control
- Cost of $25,000 or less for a minimum 4,000 byte memory
- Capable of standard computer functions (input, transfer, store, process, output)
- Operations in a normal office environment without special air conditioning or extensive electrical system requirements
- Operations under program software control.

Normally a full-range of peripheral devices including card reader/punches, printers, plotters, magnetic tape drives, discs and communication interfaces are available for the minicomputer and, with the options available, system configurations can be developed to do much of the work of a large computer system.

It is our purpose in this report to review some of the major features of minicomputers as they may be applicable to the support of the EPA Bibliographic Systems under study. Because of the broad range of minicomputers now available on the market with varying characteristics, costs and capabilities, the study could not, within the time available, address specific configurations to each individual Bibliographic System requirement.

Rather, within this report, we present some major considerations which should be addressed if EPA determines that minicomputer support is a viable alternative to current Bibliographic Systems computer processing support. Discussed here are:

- Advantages of minicomputers
- Disadvantages of minicomputers
- Planning considerations.

It is specifically intended in this report to present considerations without bias or recommendations. The feasibility of minicomputers as an alternative to current support is addressed in the ALTERNATIVES AND ACTION PLAN report.

1. ADVANTAGES OF MINICOMPUTERS

A minicomputer configuration, well planned and conceived, can offer advantages over shared use of a large scale computer facility even when such a facility is readily available.

Properly controlled, the minicomputer can offer distinct cost advantages over time shared operations when these operations exceed the cost of minicomputer rental (normally $1,000 to $3,000 per month). Care must be taken in controlling minicomputer peripheral expansion, as the cost of peripherals are not significantly less than those for standard computers, even though main frame costs remain attractively low.

Of major importance to the minicomputer user is ownership. The system is his, configured for his application and normally outside the control of the central organization. Considerations such as vying for shared facilities, special provisions for data and program security and queueing with all other users for centralized application program development are not factors in minicomputer usage.

A third major advantage of the minicomputer, provided that it does not grow to an uncontrollable configuration, is its portability and lack of need for special facilities. Standard electrical wiring and special air conditioning are not required, nor is raised flooring or a special room. Most available configurations are small and fit within the normal office layout.

Finally, depending on the vendor and the software he provides, applications development may be greatly facilitated. Well entrenched vendors offer a good, canned operating system, high level compilers and parametric coded communications and data management routines which do not require extensive EDP experience to use. This is a double sided advantage, however, as poor vendor software support can cause the minicomputer to lose its cost advantage and can damage or destroy expected application system results.

2. DISADVANTAGES OF THE MINICOMPUTER

The two primary disadvantages of the minicomputer are growth beyond its limited capabilities and poor vendor support. In addition, there are some technical disadvantages as compared to sharing a larger system which are discussed later.

(a) Controlled Growth

A minicomputer is limited in size, speed and input/output capability. Configurations are normally specified with an intended application in mind. However, like many computer facilities, a minicomputer configuration once implem ented can be called upon to perform new applications for which it was not planned, simply because it is available. While adequate for its intended application, as production requirements grow, it becomes less than adequate for all applications.

This disadvantage may be overcome by careful control of minicomputer usage.

(b) Vendor Support

Unlike its larger brother, the minicomputer can be constructed with little initial investment and minimum capital outlay. Minicomputer technology is known and well published and circuitry is available from a host of suppliers as are peripheral devices. These factors have created a mixed vendor market where, generally, the equipment is reliable and installation support ranges from normal computer industry standards to non-existent.

Thus, depending on the vendor, the minicomputer user may face:

- Unmet delivery schedules because the computer he purchased was only a drawing board configuration at the time of his order
- Non-existent customer engineering support
- Non-existent or inoperable software support or software releases which are not compatible between versions.

In addition to these problems, because of the number of vendors from which one can choose, the user may be faced with owning the machine of a vendor who has gone out of business.

(c) Some Technical Restrictions

As named, the minicomputer has many of the same capabilities as its large scale counterpart on a smaller or slower scale. Some restrictions include:

- Memory size - normally restricted to a maximum of 64 K bytes which precludes the application of large models or table programs
- Cycle time - normally measured in nanoseconds, however, often more fetches are required to complete an instruction because control memory is smaller
- Input/Output channels - normally slower and fewer available
- System redundancy - failsafe operations are usually not provided for. Of special concern is peripheral device dependency where the failure of one peripheral can down the entire system
- Software (Operating System) - the state of the operating system art in the minicomputer industry is varied depending on manufacturer and is not at the level of the standard computer. Attempts at sophisticated software have generally failed because of memory restrictions
- Software (compilers) - normally manufacturers offer a subset of the standard compiler languages (COBOL, BASIC, Assembly, RPG, FORTRAN). However, these are subsets, not compatible across vendors and often not compatible across releases within vendor.

The disadvantages of the minicomputer are often blurred by the cost advantage and the apparent need of the user to control his own computing capability. The next section, PLANNING CONSIDERATIONS, presents some factors which should weigh both advantages and disadvantages to develop a reasoned approach to minicomputer support.

3. PLANNING CONSIDERATIONS

With some of the advantages and disadvantages of minicomputers in mind, proper planning for a minicomputer can be made so that it may be evaluated against other support alternatives. The factors presented here are

- The Configuration
- Vendor Support
- Development and Maintenance of Applications.

(a) The Configuration

In order to develop a satisfactory hardware configuration, identification of the necessary hardware must be made. This must be done independently for each computer system. Eight thousand words of memory on one computer does not necessarily equal eight thousand words on another. Nor can the assumption be made that a five million byte disc provides equal usable data storage on a different computer system. There are many options available with a standard CPU and should include:

- Console panel
- Power fail with automatic restart
- Direct memory access
- Bootstrap loader
- Hardware multiply/divide
- Programs in read-only memory.

Peripheral devices must be compared on the basis of available performance not rated performance (disc storage available equals total disc capacity less the storage required by operating system software), maximum data transfer rates via direct memory access will vary greatly among different computers and, in some, may even require a second CPU or second memory part before they provide adequate direct memory access data transfer rates.

An additional consideration for the capabilities of a minicomputer must be the use to which it will be put in terms of peripheral devices. Minicomputer peripherals are divided into four general categories: a) conventional slow speed input/output devices such as typewriter, paper tape readers and punches, card readers and punches, and line printers; b) storage input/output devices such as casette tapes, small standard magnetic tape units, magnetic tape drives, fixed head disc, movable head disc drives, disc pack or cartridge drives; c) special purpose such as cathode ray tube displays, plotters, and digital interfaces; d) data communication interfaces and terminals such as single line interfaces, standard communication lines, subsystems or multi-plexers for servicing multiple lines, teletype or other keyboard terminals, and display terminals.

Every minicomputer system will not support all of these devices although most will support most devices in varying degrees. For example, a system extensively used or marketed for data communications would have many more data communications interfaces than a system marketed exclusively for process control. Business systems are usually aimed at handling a wide variety of applications and normally do not limit themselves to one specific type of interface. In so doing, they may allow, for example, hooking up to twelve remote terminals. However, in actuality, because system architecture is designed to provide a wide variety of devices, only three of four may be able to be used without substantial system degradation.

Peripheral devices represent a significant part of the minicomputer system not only in terms of system cost but also in terms of performance. Most minicomputer manufacturers do not make their own peripherals but buy them and then develop the interface and control. Peripherals are still expensive in relation to the basic cost of the processor and memory.

(b) Vendor Support

Adequate service from the vendor must consider presales benchmarking, preinstallation planning and training, post-installation technical support and customer engineering support. This can become complicated if a mixed vendor environment is selected.

Each of these support factors should be detailed with the vendor and the cost of having or not having each carefully determined.

- The benchmark - this is an essential support requirement because it offers a method of comparison and more importantly, it gives proof that the vendor configuration is functioning

- Preinstallation support - should include operations, software and programming training. The new minicomputer user normally does not have the same programming support as the large scale user and his staff must be trained for some program development

- Installation support - includes operator training, new release software training and continued software development

- Customer engineer - includes maintenance and hardware problem solving. It can be provided on contract from a service vendor or by the manufacturer and its need is dependent on the users need to have continued operations.

Each of these support factors have a cost and each must be carefully planned.

(c) Development and Maintenance of Applications

The software capabilities of most minicomputers are less advanced than those for large scale systems. Thus, complicated applications, made simpler by specialized handling routines on a standard computer, are difficult on the minicomputer. Further, because the user of a minicomputer is most often not in the EDP applications business his level of sophistication is lower.

He must consider the following to support an application:

- The need for a programming staff able to develop and maintain his application
- The lack of sophisticated software diagnostic and debugging aids on the minicomputer which may cause longer development time
- Applications development costs which exceed his original hardware investment price. Users must be made aware that the basic quotation for a minicomputer does not include literally the same requirements that are necessary on a large scale system. Programming, systems analysis, testing, conversion and many similar functions are required regardless of the size of the system.

Thus, the cost of a minicomputer consists of many components and the low hardware price which initially attracts the potential user becomes inflated by a number of factors.

4. SUMMARY

Generally speaking, minicomputers can support bibliographic systems. However, the limitations and capabilities of each minicomputer must be examined. It is beyond the scope of this report to evaluate the multitude minicomputers on the market to see whether or not they can support, in a cost effective basis, bibliographic systems. The basic minicomputer systems offer considerable processing power for the dollar but the price of peripherals is not in keeping with processor prices. Thus, the addition of a few peripherals can increase processor costs exorbitantly, raising the overall system price to a characteristically unminicomputer level. In addition, software development costs are as high as for a mid- or large-sized computer.

If, after due consideration, EPA determines that a potential alternative is supporting any or all of its bibliographic systems with program management controlled minicomputers, the factors which must be weighed in that consideration are outlined herein.